Informatica

An International Journal of Computing and Informatics

Special Issue:

Internet Based Tools in
Support of Business IS

Kevin Kelly — An Interview

1977

Informatica

An International Journal of Computing and Informatics

Internet Based Tools in Support of Business Information Systems

Witold Abramowicz
Department of Information Technology,
University of Economics, Poznań, Poland
abramowi@novci1.ae.poznan.pl
AND
Marcin Paprzycki
Department of Computer Science and Statistics
University of Southern Mississippi, Hattiesburg, MS 39604, USA
m.paprzycki@usm.edu

We have reached the point when to say that Internet has a very strong effect on the way that society operates would be a truism. In many countries, the effect of its presence starts to be felt in all domains of our existence. The remaining countries are rapidly moving in this direction. One of the areas where the changes will be especially visible is the information technology supporting business practices. It can be expected that the overall structure of how the business is done and how the companies operate will change. The early stages of such changes can be already observed. They coincide with the changes occurring in the world's sociopolitical structure. The downfall of communism reintroduced to the world the economies of Central European countries. These economies are being reformed at a very rapid pace and a part of these changes is the introduction of the information system support for business. In 1997 in Poznań the first Business Information Systems conference took place. The aim of this conference was to provide a forum for exchange of ideas about the development, implementation, exploitation and improvement of Information Technology systems in business processes.

As a part of this meeting a number of papers related to the role of the Internet in Business Systems have been presented (and published in the Conference Proceedings [1]). This Special Issue on *Internet Based Tools in Support of Business Information Systems* consists of eight invited papers. These are substantially modified and expanded versions of original presentations addressing the effect of Internet (and Intranets) on the Business Information Systems

The first paper looks from the global perspective at the effects of the Internet on the structure of business practices. Jerzy Kisielnicki, in *Virtual Organization as a product of Information Society*, discusses the basic principles of the new type of a global enterprise – a virtual organization. Such an organization is created to increase profit by cooperating organizations in various locations and the Internet is used as the communication medium in support of their efforts. The basic issues related to the functioning of such an organization are presented and discussed.

The second group of papers addresses the phenomenon of Electronic Commerce – attempts to use Internet to facilitate sales of products and services. The general overview of the current situation and its future developments is presented by Rainer Thome and Heiko Schinzer in *Market Survey of Electronic Commerce*. Although most of their examples are based on the German experience they can be easily extended to other markets. The next paper is focused on the particular niche of the electronic commerce – sale of information products. Albert Endres, in *Information and Knowledge Products in the Electronic Market – the MeDoc Approach*, summarizes the experiences gathered during the development of the MeDoc project. This collaborative project attempts to create an online book and document delivery system. The results of this experiment shed light on the future development of the information commerce. Finally, one of the most important issues that still limit the spread of electronic commerce is the widely-perceived lack of security of the electronic transactions. These concerns are addressed by Janusz Stokłosa in *Cryptography and Electronic Payment Systems*. In this overview-type paper, he discusses the electronic payment methods, cryptographic mechanisms, digital signatures as well as financial and cryptographic standards.

The next three papers are oriented toward practical applications. In *Database Support for Intranet Based Business Process Re-engineering*, Wojciech Cellary, Krzysztof Walczak and Waldemar Wieczerzycki present a new approach to the organization management based on the application of Intranet technology. Their basic focus is on the flexibility and possibility of process evolution. These two features have to be addressed to allow an effective solution to handling dynamic changes in the enterprise management procedures. Second, in *Software for Constructing and Managing Mission-Critical Applications on the Internet*, Piotr Dzikowski reviews how the application of tools like BEA Jolt and BEA TUXEDO allows fast development of the Internet applications. The flexibility, scalability and manageability issues are of special concern to the author. Similar practical concerns are

in the center of interest of Dariusz Smoczyński. In *New Software Technologies in Information Systems*, he describes how to design and implement Internet and Intranet applications for small and medium size businesses which cannot afford a large number of Information Technology personnel.

In the final paper *Communication Satellites, Personal Communication Networks and the Internet*, Hugo Moortgat presents his perspective on the development of the Internet. He argues that although currently a non-player, satellite systems will in the future play an important role in supporting both the Internet and the personal communication networks.

References

[1] W. Abramowicz (ed.) (1997) *Business Information Systems '97*, Academy of Economy, Poznań.

Virtual Organization as a Product of Information Society

Jerzy A. Kisielnicki
Faculty of Management
Warsaw University
Szturmowa Str. 3
02-678 Warsaw, Poland
E-mail: `kis@ocelot.uw.wz.edu.pl`

The aim of this paper is to present the basic principles of operating of a new type of an enterprise – virtual organization. Virtual organization is a product of the information era society. This idea is based on the free access rule. The reason for creating the virtual organization is to increase profits. The profit of individual companies operating as the virtual organization is bigger than in the case of operating separately. The virtual organizations are the result of the Information Technology (Networks, Databases). The aims, principles, activities, strategic analysis and future tendencies of the virtual organization applications are presented and discussed.

1 Aims of the virtual organization

Current civilization development is often called the post-industrial era or era of an information society. Not always one can realize that this transition is tied with many consequences, which can be positive but also dangerous.

The basic thesis which is to be to substantiated here is that the era of the information society will cause creation of completely new forms of organizations as well as new elements in the traditional ones. The new organizational forms are created in the macro scale as a new form of global and multinational organizations and in the micro scale as local organizations. These organizations are not always completely new organizations. The new forms are often created by a transformation of the already exiting ones. Such organizations are called "virtual organizations". Virtual organizations allow gaining new, until now unpredictable, profits but also bring some new threats. Opportunities and threats will be presented in a further part of this paper referring to the strategic analysis of the virtual organization.

The name "virtual organization" is a new term which is not always well defined. To define it the following questions should be answered:

- what is the virtual organization?

- is it modification of an already existing organization or is it completely new entity, which needs revision of the current terminology of the theory of management?

- what is the reason for creating a virtual organization?

The word "virtual" originates from Latin, where *vitualis* means effective and *virtus* means power [11]. The word "virtual" denominates something that may theoretically exist. Information system specialists use term "virtual memory" to name memory volume available for users disposal independently of the physical memory size.

The virtual organization represents a new type of organization, which could be created due to development of Information Technology - especially expansion of global information networks and large databases. It is also a reaction to the demands of free market and necessity of being on competitive edge [5]. The ultimate definition of a virtual organization has not been prepared yet. It can be assumed that it is a derivative of the virtual reality ideas. Certainly, it is related to the new possibilities that came from modern Information Technology.

Virtuality is described by the nature of essence, not by characteristics of physical properties. That is why one can talk about virtual organizations, virtual services, virtual travels and virtual activities. Virtualization of activities is adopted, for example, in jet pilots' training, foreign language courses and sales of tourist services. Thanks to virtualization, training is cheaper and less time consuming. In the literature of the subject the following descriptions of the virtual organizations can be found:

A temporary network of independent companies – suppliers, customers, even previous competi-

tors – connected by means of the Information Technology, in order to share skills and costs of access to new markets [2].

An artificial entity, which because of its highest utility for the client and based on individual basic competence realizes integration of independent enterprises in the processes (chain processes) of the product creation, where there are not necessary any additional investments on coordination and not diminishing the importance of the client due to its virtuality. [10].

Generally, every author presents the subject in individual manner. Therefore, every definition, including the above ones, can be questioned. Probably, in the first definition the aim of the virtual organization narrowed too much the idea of these organizations. In addition, the description "temporary network of independent firms" is hardly acceptable. If any organizations co-operate in particular area, their independence in functioning is certainly limited.

As far as the second definition is concerned, the phrase "artificial entity" is not precisely defined itself and the statement that they do not require any additional expenses for co-ordination is not always true. That is why it is assumed that the virtual organization is based on voluntary access of its members, who enter into a different new kind of relations in order to get more profit than in the case of traditional operations. It is not necessary to get any formal agreement prior to performing activities together. The duration of such co-operation is defined by this organization, which first decides to terminate it as being no longer profitable. Without this member, the virtual co- operation can exist further if the other members decide for this. In addition, new members can extend this organization [5,6].

It is assumed that the virtual organization can not be defined only by classical theory of organization. It is not so obvious that the virtual organization possesses all attributes necessary to classify it as an organization. The theory says that the organization can be separated if there is peculiarity of aims, formal structure and a preservation of knowledge. The virtual organization fulfils the basic feature; the peculiarity of aims. The existence of the two remaining features can be questioned. The virtual organization continuously transforms. It is connected with other companies, not necessarily virtual ones. Its cardinal characteristic is an adoption to new requirements, due to changes in the external environment. The organization maintains relationship with virtual organization as long as it is effective. Which means that the members of the organization are convinced that it is more profitable to remain in the organization then to be outside. The virtual organization can operate in every place where the profits are expected. Therefore, the profit is the aim of

this type of an organization. It is strongly emphasized in the literature that in the virtual organization all members are connected and nobody is discriminated or favored. The virtual organization shortens the distribution and decision making processes, which as the result brings reasonable profits to the connected companies.[1]

When the management system topology is considered the virtual organization can be, to some extent, compared with the management system of an amorphous organization. Access to such organization is free but some rules are enforced. Withdrawal is not limited, either. Each member of such an organization operates on its own account as well as a part of the organization. Figure 1 shows important characteristics of the virtual organization – it can be a member of many other virtual organizations but also can operate as an independent organization functioning according to e.g. commercial law.

The following criteria allow one to distinguish the general types of virtual organizations:

– scope of activities: single line or multiple lines

– range of influence: local, national or global organization

– clients: organizations working for the specialized clients or the general clients.

Organizations, which co-operated in the virtual mode for a long time, are often willing to formalize it and to set up a legal alliance. In such a case the virtual organization mutates into a traditional one. From client's perspective a virtual organization is satisfactory when it operates in a similar way and as well as a traditional one. As it will be shown later, the client should not be aware that his order is executed by a virtual organization. She focuses on the results only: e.g. a properly completed order. As it was said before, the organization establishes virtual relationship only when the profits are bigger than the profits achieved from traditional operating.

2   Information technology as integral part of the virtual organization

Information technology enables the creation of virtual organizations. It is a basic management infrastructure for these organizations. There are two major elements in this infrastructure: global computer networks and large, distributed databases.

The World Economy Summit, which took place in Davos (Switzerland) in February 1997, had as its motto "Building network society". Such network relations are showed in Figure 2. They can be elements of
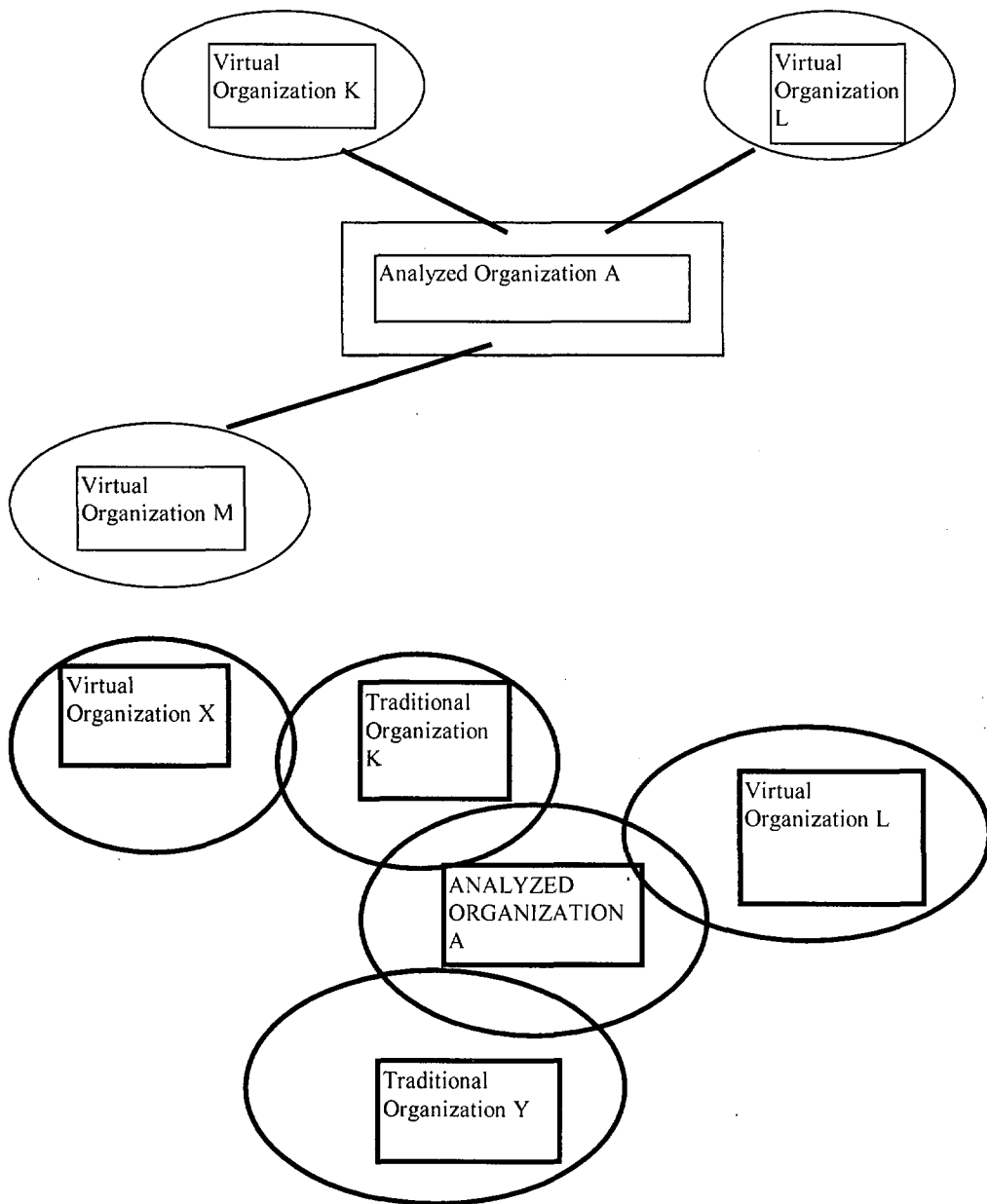
Figure 1: Examples of relationships between the traditional and virtual organization.

global networks like the Internet or some smaller network, for example limited to one virtual organization (e.g. and Intranet). It can also be a network of a WAN type or metropolitan area network of type MAN. Computer network is a medium of communication between different virtual elements of organization.

The success of any firm depends on its access to the information and its ability to utilize it efficiently. Information Technology changes the character of communication processes between the partners. It is no longer traditional, face- to-face contact between the sales person and the customer. Right now, it is an electronic or virtual interaction. For example, a Warsaw based client may use services provided by an organization located in Warsaw to buy a property in Spain. Organization, which gets the order, can execute it in traditional way or can set up the virtual organization. Thanks to such organization the client can have a look on the chosen house and neighborhood or even make transaction without leaving Warsaw. The procedure is as follows: client places the order for buying a particular property in any organization, which provide such service. Looking for an offer the local organization can create a virtual organization with a similar Spanish company. The alliance can be set up to perform only one, particular transaction or can remain longer. Effects can be observed in all the parties involved. Client gets what he wanted without spending any additional money for travels. Organization fulfilled client's requirements and receives the provision without keeping a representative abroad. The Spanish organization gains new clients and markets without any significant investments. This transaction would impossible without involving elements of Information Technology, which allows transmitting multimedia, like voice and video, simultaneously with the regular data.

Additional, virtually connected organizations can extend number of parties involved in the transaction. For example, these could be banks or notary office, which service the deal form financial and legal side. The scheme of such relations is presented in Figure 2. Client sees the complex virtual organization as one unit, despite of serious differences between each member. Furthermore, he or she is not interesting whether the another client's transaction is performed by the same or slightly different virtual organization. In reality, the situation can be even more complicated. For example the client is going to buy a property at the Mediterranean or Aegean See seaside. The client specifies requirements and declares that does not have any preferences about country, where the property is located in. In such case territory to search through increases. Organization intends to satisfy client's order should establish virtual relationships not only with the Spanish organizations but also with the Greek, Italian, and French, ones. The client, at the same time, can make use of legal or financial consulting services. The

advising organization can represent the client in contacts with the searching organization. To execute the order according to Spanish law the advisor can create virtual relationship with a local notary office. The following groups can be involved in this transaction, depends on chosen variant:

- group of independent virtual organizations,

- team consist of virtual and traditional organizations,

- number of traditional organizations.

It is assumed that the execution of the order by the group of virtual organizations is the most profitable way. However, before making final decision a comparable economic calculation should be performed.

The significant support for the virtual organizations come from the Data Warehouses. These new organizations are told to be "children" of the Information Technology. That confirms W. Titze's statement: "At the end everyone who manages 10 billions dollars using six computers is an virtual entrepreneur" [13]. Why six is needed is not said, but obviously, control is possible only if information is available, especially rare information.

3 Strategic analysis of virtual organization — anticipated directions of expansion

There can be made a hypothesis that virtualization is a way to increase competitiveness of the organization. In order to define direction of the future expansion of the virtual organizations the following strategic analysis models were used: market forces analysis of M.E. Porter [8] and SWOT analysis [12].

3.1 M.E. Porter's strategic analysis.

In the Porter's conception the organizational analysis is based on analysis of the five factors, which define its attractiveness:

- level of competition between the organizations offering similar goods or services,

- influence of suppliers and possibility of imposing a pressure on the organizations,

- influence of customers and their pressure on organizations,

- risk of introducing new, competitive products to the market,

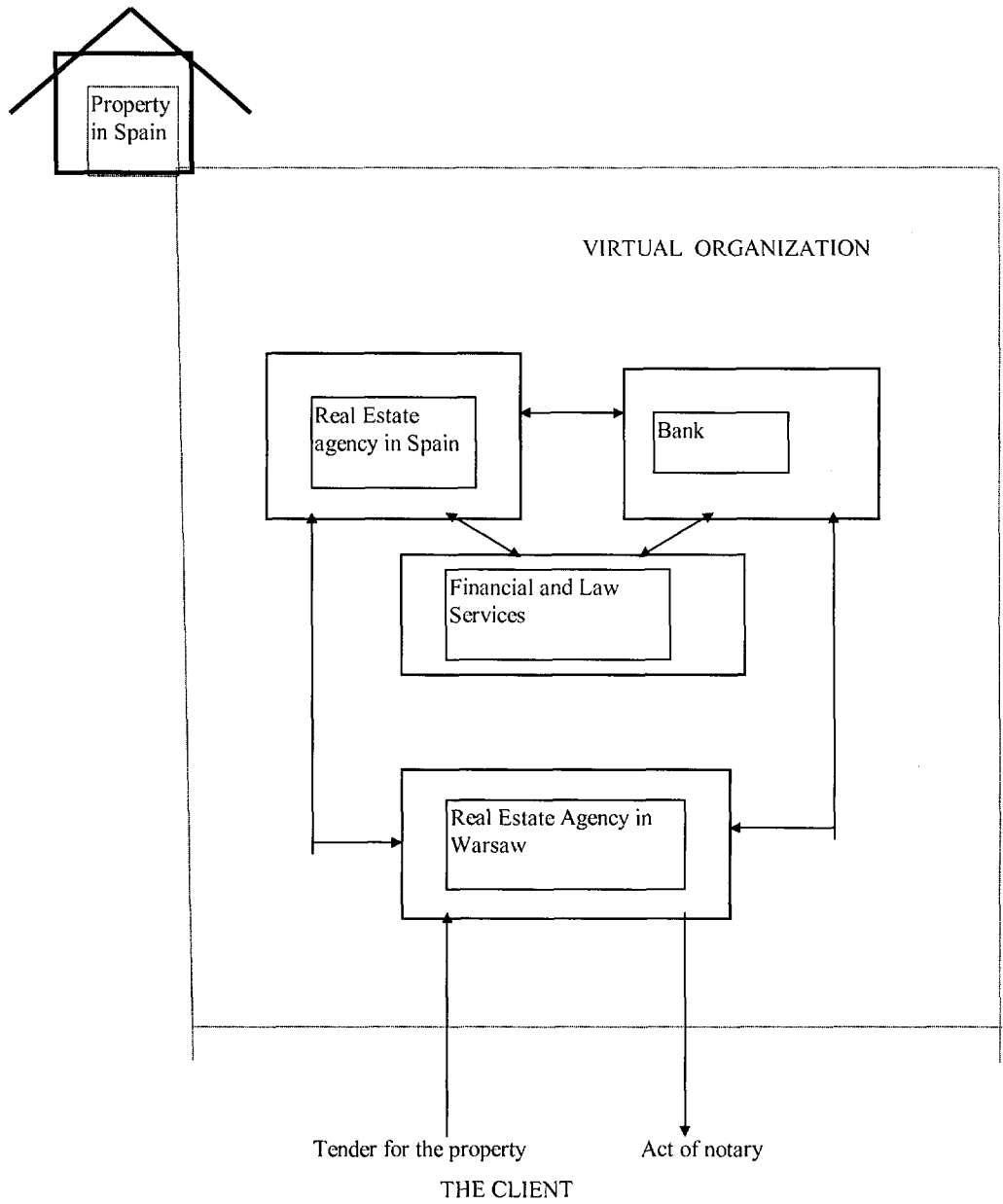- risk of introducing substitution products to the market.

Figure 2: The diagram of connections in hypothetical real estate virtual organization.

Relations with the organizations offering similar products or services allow creating new alliances. In such alliance, even the current competitors can find a place. The alliance lasts as long as is profitable for the parties. That relationship gives a chance to show the negotiation and partnership skills. Suppliers and customers are no longer on the opposite sides but they create alliances to keep common policy e.g. pricing or quality policy (the second and the third Porter's factors). It is very convenient for the organization and does not require time consuming and burdensome negotiations. Virtual relationships allow for early identification of threats due to introduction of new products or substitutes (the fourth and the fifth Porter's factors).

3.2   SWOT analysis

The analysis includes assessment of the current organizational and environment situation and anticipation of its change in the future. During analysis, the following elements are investigated: Strengths, Weaknesses, Opportunities and Threats. Table 1 shows the analysis of these characteristics prepared for a hypothetical virtual organization. In this table some selected elements of the estimation of the virtual organization are shown. To analyze a particular organization the table may be more detailed and has to be prepared especially for the given organization.

Sometimes some opportunities can become threats under special conditions. For example, Internet is seen as a huge opportunity for expansion of virtual organizations. In fact, this global network allows organization to operate on countrywide or Europe-wide base. Theoretically, it is possible to extend operation area into Australia, New Zealand, Africa and so on. Unfortunately, according to the professional magazines there are more and more problems with the Internet. The most frequent issues mentioned are long connecting time, especially in peak hours, and data security. Telecommunication operators say that the communication traffic stops telephone nets. Therefore, additional significant investments have to be made in order to increase the quality of the infrastructure to the level necessary for proper functioning of the virtual organizations.

In spite of the legal sanctions hackers see cracking the database as a challenge. There is a lot of news how fast the database security precautions are broken, even in the governmental agencies. Without solving these problems, it is hard to expect rapid development of virtual organization, especially in financial area.

The nearest future will show whether the market accepts the virtual organizations. The question is if such organizations will be created and appropriate formal and law regulations will be elaborated. Although presented strategic analysis was prepared for hypothetical

organization it seems, that it is an attractive organizational solution.

Everyone, who has the necessary technical equipment - it means a multimedia computer and an access to the computer network - and is able to use the information resources, can enter the virtual organization. This is one of the easiest and the cheapest ways of creating private organization. The legal issue is to create a suitable definition of the virtual organization in commercial law and to prepare such regulations, which will guarantee equal right for the participants and profits sharing proportionally to their involvement to the business.

The virtual organization operates as "transparent company", which means that all actions can be observed by competitors. This way the organization is forced to operate in a responsible manner. Its image is affected by the actions performed by other virtually connected organizations. Moreover, there is no common administration, buildings or properties, neither a common supervision nor control. Because virtual organizations influence each other only in a limited way, the aspect of trust between companies should be emphasized and special procedures should be undertaken to develop it.

There is not any doubt that virtualization can increase competitiveness of the organization. However, it should be remembered that management of such organization needs new, techniques, different than traditional. This is a separate problem which is only mentioned here. For example, there should be verified procedures of negotiation, which run immediately without the initial stages described in books as "Getting to yes". In the future, the virtual organization is going to be an interesting alternative for currently functioning organizations. The perspective of achieving high level of operational stability by a virtual organization, which does not have any formal structure seems very interesting and gives a chance to get more competitive advantage on the market. These subjects are very interesting themes for interdisciplinary researches, including professionals from economy, sociology, management or law areas.

4   Application

Development of the virtual organization is highly related to development of the Information Technology. Apart from new solutions available in the future, it is possible now to point to those domains of activities where application of the virtual organization should bring the most significant effects. We can define the following areas:

- different type of trades, like real estate sales, foods wholesaling and others,

Strengths	Weakness
1. High flexibility of operation.	1. Necessity of heaving financial resources
2. High speed of transaction.	required for development of Information
3. Common policy of organizational activities.	Technology, which include:
4. Transaction costs reduction.	a/ global networks,
5. Reduction of investments necessary for	b/ large databases.
organizational development.	2. Necessity of trust between all members of
6. Minimizing amount of legal services necessary to	the virtual organization.
perform transactions.	3. Possibility of joining incompetent or
7. Bringing into virtual organization the best	unreliable organization.
competencies of each partner.	4. Difficulties in collecting fees from the party
	that caused failure
	5. Lack of formal supervision and possible
	difficulties in co-ordination of transaction
	execution.
Opportunities	Threats
1. Fast reactions on changes in the environment,	1. Technical equipment inefficiency noticed in
especially on the emerging market niche.	networks e.g. In the Internet, where full
2. Execution of transaction despite of any law,	multimedia transmission is hardly available.
organizational and other barriers.	2. Lack of law regulations in respect of
3. Possibility of utilizing the up-to-date methods and	operations and responsibility to the members
management techniques.	and clients of virtual organizations.
4. Possibility of co-operation between parties, which	3. Lack of clients' and companies' readiness
in other conditions would never co-operate.	to use virtual organization services.
5. The information alliances operate over custom	4. Psychological resistance, caused by
borders.	changes in the way of performing
6. Reaching new group of clients.	transactions.
7. Lack of cultural, race and other prejudice	
between parties of the transaction.	

Table 1: SWOT analysis for a hypothetical virtual organization.

- sales of tourist services for individuals or tour operators,

- different kind of services like banking services.

These three basic application areas seem to be the most effective for the virtual organizations. However, the range of possible applications can be much wider including consulting or designing information systems.

Right now, it is difficult to show operating virtual organizations. The company which can be regarded as a virtual one is the company which uses electronic money - Ecash. The system has been design by Dutch company; DigiCash [9]. It is completely different concept than the idea of electronic cards, because it reverses cash payments in the real world. Digital notes that, as in reality, have value, unique serial number and safeguards against falsifying are used in the transactions. Traditional banks like the Mark Twain Bank in United States and Merita Bank in Finland are the central points of the system. Banks like Deutsche Bank and Advance Bank form Australia is going to join the system. Client, using special software withdraws some amount of money, which is "deposited"

on PCs hard disk in the form of digital records. After that, the client can do shopping in virtual or traditional stores.

A lot of tourist agencies and personnel agencies as well as consulting companies present some characteristics, which are unique for virtual organizations. They also advertise themselves as such organizations. One of such examples may be travel agency "Rosenbluth." The owners of the agency decided that developing the company in a traditional way, which meant buying the existing travel agencies or creating new ones all around the world was not acceptable for them because of the costs and time. So, they created a network of 1300 offices in 40 countries using virtual organization principles.[3]

Sometimes, organizations like mailing bookstores consider themselves as virtual organizations. It is rather a marketing slogan, because the management usually assumes that client should associate such new phrase with the most modern level of services. However, to classify a bookstore as a virtual organization many conditions have to be fulfilled. Client should have a chance to choose a book almost like in a real

bookstore.

We can make a hypothesis that the development of the Information Technology will cause rapid increase in the number of the virtual organizations. This statement can be considered true because SWOT strategic analysis of the virtual organization indicates advantage of the S.O. factors over the W.T factors. It means there is supremacy of strong sides of organization and its opportunities in the environment over weaknesses and threats. Creating virtual organization or taking part in a system of such organization it should be remembered that apart from profits there is possibility of loses, especially in situations which are not under our direct control.

References

[1] Benjamin R., Wingand R. (1995) Electronic Markets and Virtual Value Chains on the Information Superhighway, Sloan Management Review, 2, p. 62.

[2] Byrne J.A., Brandt R. (1993) The Virtual Corporation, Business Week, February, 8th.

[3] Clemons E.K., Row M.C., Miller D.B. (1992) Rosenbluth International Alliance: Information Technology and The Global Virtual Corporation, (eds.) Naunamaker jr. J.F., Sprague jr. R.H., *Proceedings of the 25th Hawaii International Conference on System Sciences*, vol. IV, Los Alamitos, CA, p. 678.

[4] Holland C., Lockett G. (1995) The Creation of Virtual Global Banking Partnerships, *Proceedings of the 3rd European Conference on Information System*, Athens, p.1111.

[5] Kisielnicki J. (1997) Wirtualna organizacja, marzenie czy rzeczywistosc, Computerland, 16, p. 58.

[6] Kisielnicki J. (1997) Wirtualna organizacja jako organizacja przyszlosci, *Wiosenna Szkola Polskiego Towarzystwa Ekonomicznego.*

[7] Kisielnicki J., Parys T. (1997) System informacyjny w wirtualnej organizacji, PC Kurier, 7, p. 17.

[8] Porter M.E. (1994) Strategia konkurencji, *Metody analizy sektorów i konkurentów*, PWE, Warszawawa.

[9] Rafa J. (1996), Ecash - narodziny wirtualnego pieniadza, Netforum, 1.

[10] Scholzch Ch. (1996) Virtuelle Unterhmen - organisatorische Revolution mit strtegischer Implikation, Management & Computer, 1, p.27.

[11] Tokarski J. (ed.) (1981) *Slownik wyrazow obcych*, PWN, Warszawa.

[12] Thompson A.A., Strickland A.J. (1997) *Strategic Management, Concepts and Cases*, Irwin, Boston, MA.

[13] Titze W. (1996) see Hoffmann W., Hanebeck Ch., Sheer A.W., Kooperationsborse - Der Weg zum virtuellen Unternehmen, Management & Computer, 1, p.35.

Market Survey of Electronic Commerce

Rainer Thome and Heiko Schinzer
Department of Business Computing
University of Wuerzburg
Wuerzburg, Germany
E-mail: thome@wiinf.uni-wuerzburg.de

The paper presents and overview of the most important issues related to the Internet-based electronic commerce. The market potential and scenarios of participation are introduced. Role of most important service providers for the electronic commerce is discussed and illustrated using examples primarily based on the German electronic commerce market.

1 Introduction

Electronic commerce (EC) facilitates the extensive, digital transaction of business processes of enterprises and customers via the global public and private nets (Internet). It is going to change fundamentally these business transactions between the enterprises and customers. Similar to the introduction of a complex standard software, an extensive EC solution strongly influences the present organization of the enterprises and enforces an integrated information processing in order to facilitate shorter handling and process periods. Partly, EC will also influence the economical structure. Parts of business can be substituted or completed by a direct digital distribution. The resulting market potential of the EC is enormous and the expectations in the changes of the economic cycles made by the EC are so important that the American President and the European ministers deal with this topic in their own initiatives. This article gives an introducing survey of the EC market. Therefore market previsions are discussed which show the expectations in electronic commerce. Additionally, different participant scenarios are described to answer the question who can do digital business with whom. Finally we will have a closer look at the different providers of the electronic commerce. This group entails providers of hardware and software, Internet providers, telecommunication enterprises and providers of on-line services.

2 Market potential of electronic commerce

The Internet as the worldwide communication platform, offers a multitude of services like the Electronic Mail (e-mail), World Wide Web (WWW) or the File Transfer Protocol (FTP) and does not only represent a considerable part of EC but it also represents a large part of the infrastructural precondition for the EC and the border-crossing economizing of the electronic market place. The Internet with its services (e. g. e-mail), however, may not be equated with the electronic commerce. Electronic commerce concentrates on business management and organizational problem solutions whereas the Internet has its main emphasis on the technology, detached from the concrete applications.

Globalization of markets, diminution of periods of product development and production times together with the reduction of the manufacturing depth and intensification of the collaboration of cooperating enterprises indicate the present situation of the economy. An increasing decentralization of the market participants together with a growing cooperation dynamic requires new forms of cooperation between enterprises as well as new interaction instruments between the enterprises and customers.

Management attempts like virtual enterprises, business process reengineering, lean-production, continuous system engineering as well as the globalization of obtaining and selling markets or efficient consumer response are based on innovative information and communication technologies. With the help of these technologies enormous competition and development potential can be realized.

Some years ago, networks and computers were reserved to few specialists, but today they are already an integral part of the everyday life. The multimedial processing of information, the simple handling of the applications as well as the global nets induce enterprises to use modern information and communication

technology (I&K), which was used in the past in an insufficient way. Curiosity and the enormous possibilities of cyberspace lead to a change in values in the society. In the beginning, the Internet was rather an in- phenomenon and therefore it was interesting, but now more and more private households use the possibilities of a worldwide information platform, which increasingly can easily be applied for serious tasks and business transactions. Electronic commerce summarizes the use of different information and communication technologies and nets for electronic transactions of economic business. No single technology or technical detail are in the foreground here but rather the high level function-related, comfortable solutions which allow both, enterprises and private households, to interact better and faster in the market.

Until now, economy has not embraced the EC. The alteration of the competitive situation, as well as the change in the values in the working world and society, however, force enterprises to slowly accept this new challenge. Providers have not only the possibility to develop a completely new market with enormous volume and growth rates, but they have also the task to offer problem-oriented EC- services. For this, a turn away from the traditional orientation on individual technologies (like net infrastructure, telecommunication services or EDI) is required. Problem solutions are sought after which will unite all required information and communication technologies which have integrated the directly matched attempt to the task.

Electronic commerce is the prerequisite for the creation of electronic market places where information and goods can be exchanged between single enterprises, between enterprises and private households and between private households and private households. Electronic commerce is a combination of telecommunication nets and services as well as applications in connection with business strategies with which, in different compositions, EC-solutions are developed. Therefore electronic commerce stands for the technical-organizational processes, which make the execution of business transactions faster, simpler and more efficient with the help of telecommunication. Internet and especially the World Wide Web (WWW) offer new impulses for the EC. The Internet euphoria has considerably contributed to the EC discussion and to establishment of standards and international availability providing an ideal economical infrastructure. The Internet with its different services (e.g. e- mail, WWW, FTP) is not only on its way to establish as a new medium for EC but it also gains in importance as a new way of distribution. In the year 1994 the EC sales volume was about 60 million US $. In the year 1995 it was already 600 million US $ and 1996 it was about 1 billion US $. Worldwide prognoses forecast 70 - 200 billion US $ for the year 2000. Because of this, doing business, particularly marketing and sale, via the

Internet is getting more and more attractive. Furthermore it offers good prerequisites to support further EC-application fields like providing information for clients, advice and support, and it completes the existing methods of the inner and inter-company flow of information (e.g. EDI, electronic payment transactions). This becomes also clear in the forecasts made by worldwide renowned market research enterprises like Frost & Sullivan GfK, OVUM, BIS, DATAMONITOR or Diebold. With a transaction volume of more than 1.1 billion US $ (estimation DATAMONITOR) for the year 2000 in Europe and the tendency to increase, the field of EC is said to be one of the most interesting markets. For the year 2001 Frost & Sullivan (9/95) stipulate that the EC-market volume will increase up to 3.2 billion US $ (1994, 700 million US $). Diebold [2] forecasts for the EDI market, which is a classic EC service in the professional business-to-business communication, an increase rate of messages in Germany of 40-50% per year.

Isolated forecasts for the German EC-market are rather scantily available. One forecast, which is often quoted, can be found in the study Business Digital, which was carried out by Diebold and Telemedia in 1996. According to this study, the turnover of goods induced by the digital sale, in the hardware/software-industry, publishing houses/media, telecommunication, trade and tourism will rise from approx. 0.1% (corresponds to 800 million DM) in 1996 to more than 1% (9,5 billion DM) in 1998 up to 30 - 50 billion DM in the year 2000, which means a share of 3-5 % [2].

Despite skepticism towards single numbers in these analyses, the trend is clear and points upwards with high increasing rates. Among other things, the market forecasts justify themselves by a change in communication behavior between enterprises as well as private households. This is characterized by the increasing acceptance of electronic media.

What do the enterprises expect of Electronic Commerce? At first sight, amazingly they do not primarily expect quantitative effects like the rationalization and cost saving as well as increase in the sales volume and the increases in profits, but 71of their enterprise (compare Figure 1). By looking closer at the answers it can be seen, that until now the over 600 enterprises primarily have realized solutions where the presentation of product and enterprise information had priority but they still were not orientated to the digital transaction of business.

2.1   Aims of electronic commerce

The Internet, however, is not only interesting as an advertising medium for many goods and services, but it rather supports the aims of business management in their main fields like purchase, production and sale. This can be explained in part by the fact that, the cost
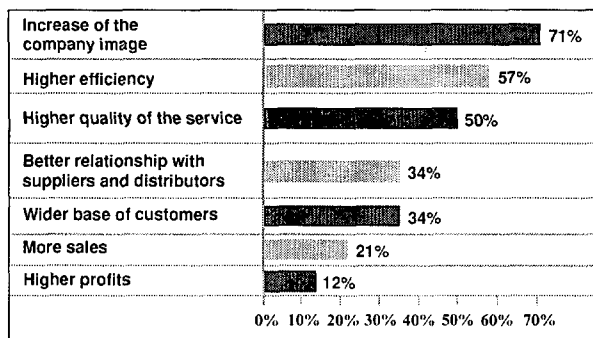
Figure 1: Survey of 603 European enterprises [1].

and therefore the prices of the products are dependent on the relatively fixed production costs and the variable sale costs (including marketing and logistic costs). Due to the extensive rationalization in the past years, for many enterprises the potential of cost reduction is largely exhausted in the field of production costs. The high amount of variable costs, however, which results from the construction of full-coverage marketing structures, can be reduced by using extensive EC-offerings (if the capacity of the sale is completed adequately or partly replaced by EC).

If the cost reduction by EC turns out well, then the profit of sales via the Internet, which were only marginal until now due to the high production costs, will rise considerably. In particular the sales market for the SMEs can be maximized by EC, because the products can now also be offered in regions and countries where the construction of classic sale and logistic structures had not been rewarding. Thus with decreasing sales costs and increasing trade, an altogether very positive profit effect for the enterprises will be the case. Due to the receipt of a part of the sale costs (line charges etc.) and the possible loss of the "buying feeling" it has to be taken into account that an adequate reduction in the price shall be given to the customer as a kind of incentive.

Corresponding to the existing EC-services like EDI, WWW or e-mail the effects of use are very different in the individual cases. Essentially the obtainable potential of use can be subdivided into the following fields:

- reduction of costs by means of information logistic,

- speeding-up and intensifying of the flows of information,

- provision of a problem-oriented information logistic,

- raise of marketing performance, and

- improvement of the organization of the building-up and flowing-process.

2.2 Cost reduction by means of information logistic

The example of EDI clearly shows, that in the past information and communication technologies were used only as an instrument for the automatization of routine transactions. If many transactions have been carried through, it could have also resulted in considerable savings due to the reduction of recording and administration costs of the exchanged business data. Only rather slowly, however, the businesspeople are realizing, that the use of different information and communication technologies can make inner and inter business transactions faster. Due to the fact that unimportant transactions are eliminated, these technologies can be considerably more effective and they can be designed correspondingly to a logistic chain (business reengineering).

2.3 Improvement of the flow of information

Electronic commerce improves the flow of information within the relational net of economic objects. Here, especially the fields of marketing and sale, management, transport logistic as well as research and development are to be named.

Marketing and sale are particularly supported by electronic commerce. The WWW as an Internet-service stands out because it is multimedial and it can be reached all over the world, because of its interactivity and topicality as the ideal marketing and presentation medium. Additionally the contacts with the customers can be statistically evaluated. Moreover, it is possible to address the customers directly and worldwide and the short-term dispatch of information of any kind (e.g. product information, company information, press reports) can be realized per e-mail. The low costs in comparison with classic media and the enormous reach are broader aspects which speak for this electronic medium.

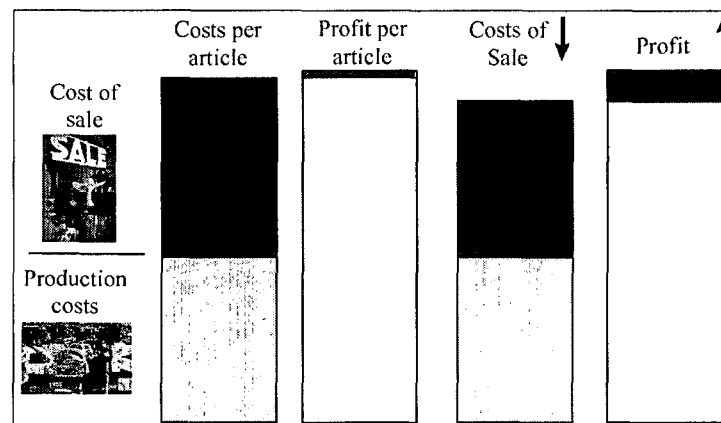Management tasks are a general field of use of elec-

Figure 2: Effects of the digital sale on the product expense and profit.

tronic commerce. Central tasks of the management are the planning, instruction, leading of the stuff, organization and control. All tasks mentioned here are based on an intensive flow of information, both within and outside an enterprise or an organization unit. Electronic commerce permits an immediate and worldwide communication with individuals (e.g. e-mail, voice-mail, conferencing, GroupWare) as well as application systems (e.g. report systems, data-warehouse).

Transport logistic is a field of application of EC with direct effects on the provision and physical movement of goods and people. In this connection EDI is a central component of the EC-concept and has a central position because it integrates business data and controls the carrying-out process. Additionally it has tracking-systems which track the sending of transports of goods; direct data link solutions, which inform the partner about the present production level of an order or the stock level of a customer. Many enterprises now have not only the possibility to sell their products on-line, but they can also deliver them the electronic way. Information suppliers like publisher (e.g. electronic publishing), suppliers of on-line databases as well as software suppliers and brokers (e.g. real estates) are to be mentioned here.

Research and development are getting completely new perspectives with application of the electronic commerce. Besides the international, economical and user-friendly information exchange between the persons via the Internet- services the like discussion forums or e-mail, software archives, patent data bases etc. can be accessed as well. Independent of the national telecommunication companies and own telecommunication services the communication medium Internet permits the development of research and development communities in the sense of simultaneous engineering (e.g. CAD data exchange) and the border-crossing use of the workflow and workgroup-systems.

2.4 Improvement of market results

The possibility of a multimedia, worldwide electronic presence, which permits the cooperation even over big distances, as well as the carrying out business on a purely electronic basis, opens a wide spectrum of new or better market results. Thus for example information on enterprises, product and service can always be prepared immediately and multimedia presentation can be provided even to an international customer without the time delay. Due to the immediate dialogue, the customer contact can be intensified. Electronic commerce is present 24 hours a day and 7 days a week, so that the customer does not have to pay attention to the opening hours when passing orders or making inquiries. He can obtain electronically transferable goods like application software, updates, information concerning his subject, publications etc. at any time he wants.

2.5 Obtaining and improvement of the competitive situation

By means of a multitude of different strategies like JIT, virtual business, BPR, TQM etc. it is possible to manipulate various competition factors like rapidity, cost reduction, quality or strategic partnerships. These strategies in turn absolutely require an integrated flow of information and business. EC with its multitude of communication and integration possibilities forms the precondition to realize the required intensive flow of information across the department enterprise borders and to allow integrated, media-break-free data processing between the individual business partners.

3 Scenarios of participation in the electronic commerce

Who are the participants of electronic commerce? In general every person who transmits data electronically within the scope of a business transaction is a participant of the EC. In the following the existing connections between business and consumer which are on the market at present are refurbished and deepened.

3.1 Business-to-Business

The business-to-business communication makes the highest demands on the electronic commerce. Volume of trade and business safety require a guaranteed security of business and data, full-developed products, standards and the additional use of the EC services, which is not purely qualifiable but also quantifiable. The fields in which EC is applied are various and contain the sales volume oriented tasks of marketing, particularly the management of sale (e.g. progress numbers) plus the operative business management also in the field of research and development (e.g. simultaneous engineering). The required EC- technologies and EC-applications as isolated solutions are largely already available on the market.

Despite the resonance the Internet has with the WWW with the final consumer, the greatest share of the EC-market will be the inter-company business transaction. In the following figure we see the increasing use of the Internet for the digital business processes between the enterprises. The absolute number is still rather modest though, so that a further strong increase is expected also after the year 2000. The flat increase until 1997 and then the precipitous growth in 1997 can be explained by the fact that in the USA as the technological forerunner of EC the critical mass in the EC surroundings has been reached and in many enterprises EC has become a decisive instrument of competition.

Until now enterprises in Europe and Germany are hardly active in this segment except for a few companies. They still work on concepts and prototypes. Here the decisive jump is expected for the mid or end of 1998.

The application of EC is one part of the companies strategy in order to keep the competitiveness (e.g. electronic hierarchy) and/or for the extension of it. The process to decide for the use of EC is comparatively long and is aimed at a long-term use.

3.2 Business-to-Consumer

With the business-to-consumer relations, the market directed functions: marketing and sale, stand in the foreground. The simple handling of the EC- services

used (particularly the WWW) as well as the attractive multimedia processing of the offer, are most important. Because of the image reasons the electronic presence on the market is essential nowadays, whereas up to now the use and cost aspects are in the background. EC is a completely new form of the customer communication. The strategies concerning the use of EC are still developing. Till now high value is set on the optical attractiveness.

The number of potential customers is increasing rapidly worldwide. The number of households in USA and Canada which are connected to the Internet has been increasing from 15.4 million in 1996 to 38.2 million in the year 2000. This would mean that via these connections about 100 million North Americans can be reached. In Europe the growth rates are even higher since the starting standard was clearly lower. Here the number has been increasing from 3.7 to 16.5 million households (see Figure 5). Among other things the forecasted discrepancy for the year 2000 can be explained with the telecommunication costs which are too high and therefore an intensive private use is prevented.

In addition to the quantitative developments, it is very interesting, to what extent the user is prepared to do business via the Internet. In Germany, for example, there is a growing willingness to use the Internet in order to buy something and to settle the invoice on the digital way.

3.3 Communities of Interest Networks (COIN)

The idea of COIN, which was developed in the USA, is to use the Internet not only for simple purchasing processes between two participants (supplier - customer) but also to represent more complex business processes where more market participants are involved who are interested in a business transaction [5]. In the following Figure 6, the COIN-prototype of buying and selling of real estates is shown as an example. In addition to the interested persons (potential buyers), property and real estate agents, experts, mortgage banks, notary offices and land registries are united in one COIN.

If the client decides for a specific object, a virtual file is created and the COIN-system does the required steps which still have to be fulfilled until the conclusion of the sales contract. In our example it is to seek for an expert opinion from a specialist, to get an offer from the mortgage bank as well as to inspect the history of the real estate with the land registry (online). The interested person decides on the order in which the different steps of the process are to be made. It has to be decided which information from which object file he wants to make available for the individual market participants.

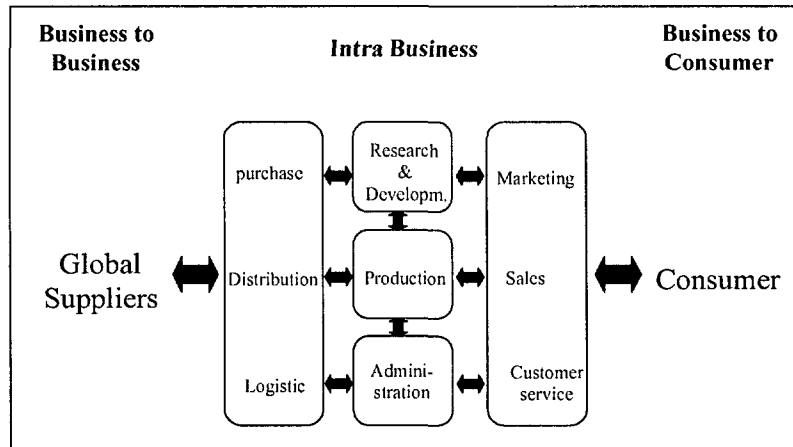Moreover the COIN system also takes over the

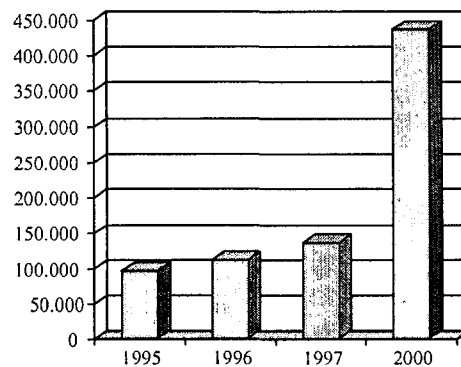Figure 3: Scenarios of the participants of Electronic Commerce [3].

Figure 4: Number of engaged enterprises for intra business (Volpe, Welty & Co.).

project pursuit, for example by reminding the individual participants to keep the date etc. With the COIN- attempt the borders between business-business and business-consumer are blurred or overcome, since both business-consumer and pure business-business-processes can occur within a business process.

3.4 Consumer-to-consumer

Till now, potential that turn out in business relations between private households and EC are hardly worked with. It is true that the use of an e-mail for the exchange of private messages has been developed enormously, but it can be counted only partly as a form of EC, because in most cases there is no economic intention. The publication of small ads in the WWW is an attempt for which a publisher takes a central mediator position.

4 Suppliers of EC-products and services

Who are the suppliers of electronic commerce and what do they offer to whom? Till now only seldom there have been many established suppliers from all fields who decided to participate in such a young market with so much expenditure. Fast agreed strategic partnerships, expensive takeovers of young technology enterprises as well as the explosion of the stock prices of hardware and software enterprises like Cisco and Netscape clearly shows that many suppliers want to create a good competition situation for themselves on the EC-market.

In order to analyze the structure of the suppliers, the following Figure 7 deals with the content of the demands. Depending on the complexity of the EC-solution, an enterprise uses an electronic product catalogue, a complete online-economic system for goods. The most extensive step is the application of pay, production and distribution software. Not only the costs of the basic software are rising but also the de-
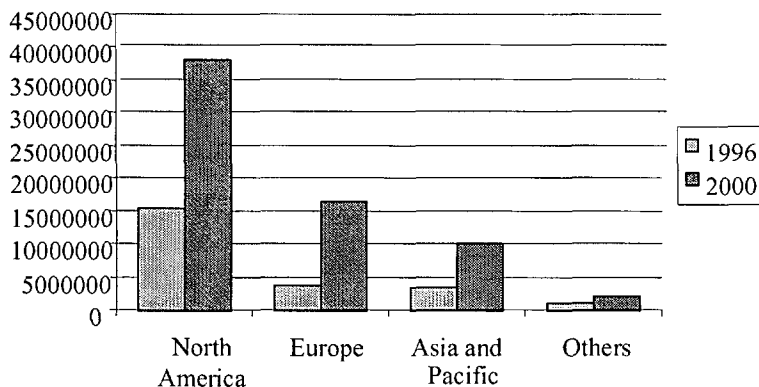
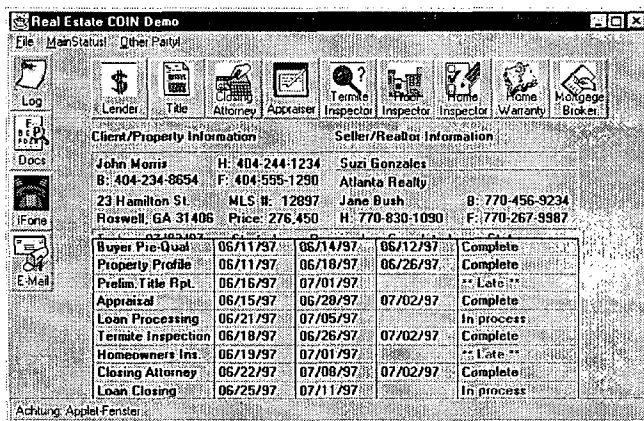Figure 5: Households with a connection to the Internet [4].

Figure 6: COIN prototype of selling real estates.

mands of the needed infrastructure are getting higher and higher. That means that the market for the EC-suppliers is composed of software for EC- clients (mostly free of charge WWW-browsers), EC-server (online-shops or professional WWW-servers), security systems (cryptography, firewalls, electronic cash) as well as hardware-components for the functioning and the security of the systems.

Due to EC a completely new segment of the Internet has been established, where all established and new founded hardware and software suppliers, classic value added service performers for the business-business-communication, Internet provider as access, content and solution-providers, telecommunication suppliers in the network field as well as established online-service providers as content providers can be found.

Almost all enterprises try extensively to establish themselves on all net product steps of an EC solution. Already in this early phase this leads to a concentration on some big EC-providers who commercialize extensive solutions by strategic partnerships with young, innovative enterprises.

4.1 Hardware and software suppliers

The established hardware and software suppliers have put together a complete product spectrum by taking different development attempts. Due to the fact that Microsoft was too late when it realized the trend to the Internet and the professionalization of EC, the start-up company Netscape gained the leading position in the consumer-surroundings of the WWW-browser. Only with the help of strong marketing and the free delivery of their browser Explorer as supplement to the operating system, Microsoft could regain and extend their share in the market.

Electronic malls are the technical information copy of the traditional shopping centers and consist of a collection of virtual, not physically present, shops which are particularly directed to the final consumer (business-to-consumer) as a new channel of sale. The counterpart for the business-to-business field are the electronic fairs. Unlike the classic shopping centers where the customer has to be physically at the location, with online-malls she has a world-wide access. In
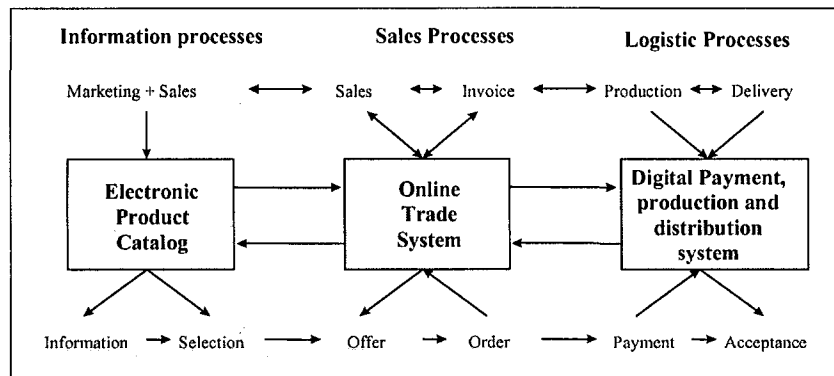
Figure 7: Survey of the spectrum of EC-products.

order to find the information or to make a transaction, the customer has the possibility to enter this shopping center by the medium WWW from an arbitrary point in the world at any time she wants.

Among the shop suppliers many small venture capital financed enterprises have established themselves as developers due to the fact that successful IT- enterprises and primarily suppliers of standard software for business management got in very late. One example is the Intershop Communication, an enterprise founded in Jena, which made licensing agreements with many enterprises and therefore it could have obtained enormous sales.

In the meantime, however, almost all top enterprises of the Internet branch like IBM, DEC, Microsoft and Tandem offer software solutions for the development and functioning of electronic market places. In the opinion of many market analysts a supersaturation is already recognizable. About 20 almost equal and relatively economical online-shops or mall-suppliers are already on the market. Therefore the late decision of SAP to found the joint venture enterprise Pandesic together with Intel, which will introduce a solution on the market in 1998, is partly judged skeptically. It will be interesting to see how the market will change for shopping systems, if all standard software suppliers like SAP, Baan, ORACLE, Peoplesoft and others permit a simple integration of own online-shops in the internal systems of the enterprises and then replace the independent online- shopping systems. Moreover, suppliers for special segments of an EC-solution are primarily represented with steep growth rates in this market, like Cisco which is the dominating enterprise for network routers.

4.2   Value added services

A value added service (VAS)-company can be an independent commercially oriented organization which

offers services in the field of EDI, as well as a network service which offers not only a standard service but also a chargeable extra service for its customers. These offered services range from the simple "EDI-To-FAX" to a completely integrated EDI business-system.

The VAS companies are specialists in the business-business-communication and know the standards of enterprises for the digital handling of business processes. The leading enterprises in the EDI-surroundings like General Electric Information Systems (GEIS), Harbinger and Sterling Commerce have discovered the Internet to put through finally the idea of EDI in a larger market. Therefore the value added services concentrate primarily on the field of business-business-solutions and work in the background, so that they are never or seldom visible to the user.

4.3   Internet-Provider

As far as just in Germany the telecommunication suppliers took a long time to recognize the market potential of an Internet-provider, a multitude of small providers has developed. In the meantime, however, the Deutsche Telekom aggressively markets its services as access-provider, so that movement toward concentration is recognizable; in order to finance high fix costs for fixed lines for broad data communication. In order to increase their chances in competition, Internet providers started more intensely to create also contents for enterprises, that means WWW-applications. Many providers, however, do not have the know-how of business management to develop or adapt EC-solutions, which go beyond the presentation of information. There is a clear danger for completely uninformed, but Internet-interested managing directors to contact such more layout-oriented providers.

4.4   Telecommunication suppliers

In the meantime telecommunication suppliers have fully realized the importance of electronic commerce. They want to realize solutions while they do not want to restrict to the classic role of the infrastructure provider. Thus, for example, the Deutsche Telekom offers not only the access to the Internet and intranet-connections, but it also offers online shops. The telecommunication suppliers use their major strength in order to obtain the partly missing know-how in the field of EDI and online-shopping by cooperation.

The proximity to the final customers as well as to the enterprises gives the telecommunication enterprises good competition chances in the EC market. After all, just in Germany the telecommunication suppliers belong to the winners of the Internet and EC-boom, since all users are forced to use the telephone lines in order to get access to the Internet.

4.5   Providers of online-services

Online-services see their classic field of use primarily in the business- consumer communication. Electronic publishing, online-shopping, information and presentation services as well as support-applications are the services which are offered in particular. In the beginning the Internet-boom has surprised online service providers like Compuserve and T-Online. While, for example, the Deutsche Telekom has strongly profited from this development with its access to the Internet and after more than 10 years it has become a very successful enterprise, Compuserve has fallen into a serious crisis because of the Internet, since many of the chargeable Compuserve-offers are now freely retrievable in the Internet free of charge. The consequences which arise from the takeover of Compuserve and AOL will be very interesting.

5   Conclusion

The idea of a worldwide network for the exchange of business data is faster explained than realized. Since the technical infrastructure in form of the Internet is available nearly worldwide, the attention must be focused on offers concerning the content. The aim is the semantic integration of the business process, which facilitates the computer of one enterprise to interpret the data of the computer of another enterprise correctly.

The realization of EC offers gives many enterprises the chance, to remove bigger quantities at lower sales costs and with it increase the profit. This big attractiveness and the global availability of the EC-solutions, however, increases the pressure of competition for those enterprises that still use the classic ways of distribution for selling their products. On the one hand many enterprises have now the chance to push

forward in regional markets in which they have not been present before, but on the other hand they will meet competitors who have been completely unknown in Germany till now.

The large number of suppliers underlines, that a enormous market has been arising. This market, however, is at first for the providers of EC-products and services. The proof, however, which shows that then there will come the success for the enterprises is still missing. Only as of 1998 this proof can be counted on.

References

[1] Gall, o.V. (1997) Survey of the EC-Marketplace, Wall Street Journal Europe, Handelsblatt, June 3rd.

[2] Diebold and Telemedia (1996) Die grosse Multimedia-Studie von Bertelsmann Telemedia und Diebold, Business Digital.

[3] Meitner, H. (1997) Electronic Commerce - Vision und Wirklichkeit, Discussion Paper presented at the 5th Multimedia Conference.

[4] Jupiter Communications (1997) Survey of the EC-Marketplace, http://www.jup.com, July 10th.

[5] Dill, J. (1997) Community Of Interest Network (COIN) Business Models, CommerceNet Research Note.

Information and Knowledge Products in the Electronic Market — The MeDoc Approach

Albert Endres
Department of Computer Science
University of Stuttgart
D-70565 Stuttgart, Germany
E-mail: Endres@informatik.uni-stuttgart.de

In the future many goods and services will be traded in the electronic market. This will include information products as carriers of scientific and technical knowledge. The electronic market has the potential to overcome some of the shortcomings of the conventional market, but will also have its own peculiarities.

The MeDoc project is exploring important aspects of the electronic market for scientific literature in computer science. A number of publishers, libraries and users are cooperating in order to propagate and to evaluate the use of books and journals as online offerings.

1 Introduction

Whoever looks at the impact of the new media on the scientific information and publication process will agree on one thing only, namely that a change is occurring [5]. It will affect everybody who is involved, namely authors, reviewers, publishers, distributors, librarians and readers (or consumers). The current debate is about the direction of the change. There are different views and different scenarios depending on the specific role of the specific player.

In this paper we will assume that goods and services offered in this field are subject to the rules of the market. This is not obvious or agreed to by everybody. In fact, this is one of the important issues: should information and knowledge products be subject to a market at all or should they be available to everybody as a public service. This issue has major sociological and political implications and is therefore of concern to national or state governments. To narrow down the topic we will make a distinction between information and knowledge, the latter to include most scientific and technical material and to exclude such areas as entertainment (or infotainment).

As a widely accepted form of an electronic market we will, of course, look at the Internet. It has some peculiarities which one has to be aware of. The project MeDoc which was initiated by the German Informatics society (GI) will be used to illustrate how the Internet can be used as an electronic market for information and knowledge products. Finally, we will give one view how the market might evolve in the future.

2 Information and knowledge as commercial entities

Very often the terms information and knowledge are used interchangeably. It is certainly helpful to look at the differences. Any message or string of symbols that can be interpreted by its receiver contains information. If not, it is considered as noise (also if it is not of an acoustic form). The information in a message may consist of several parts or pieces, and can have a superimposed structure. This structure can be quite complicated, but it is a syntactical one. The same information can exist on different media or can come in different forms. The number "seven" can either be in stone as a Roman figure (in that case as VII) or on paper as an Arabic number "7". Information can be received and processed much easier than knowledge.

Only certain types of information represent knowledge. It is that information that can be grasped by the human mind or by the members of the human society. It is information that either helps in explaining the world (orientation knowledge) or that guides our behaviour (action knowledge). It is not necessary that we understand all phenomena or sensations with respect to their reason and origin (yet). We only know that it exists and that it has certain relevance.

Knowledge is structured completely different from information. It is the content or meaning that counts. We distinguish between the broad (general) knowledge and the deep (expert) knowledge. General knowledge is what every member of every group (class or stage) of a society should have, expert knowledge is what

distinguishes one group or one person from the other. Human knowledge is such that it is impossible for everybody to know everything. Because of our mental capacity and the limited duration of our lifetimes we have to be selective.

We can be inundated with information, but not with knowledge. There is the well-known example of the LANDSAT satellite that sends about 1 Terabyte of information to earth every day, but there are not enough scientists to look at it. Only certain parts of this information will become knowledge. Knowledge needs information to articulate itself. Otherwise it can only be poorly communicated and cannot be transferred.

Economically speaking, both information and knowledge are immaterial goods. They are not traded as such, but only through their expression or materialization. As Figure 1 shows, knowledge can have many different forms of expression. Not all of them are associated with strings of text.

New knowledge is gained by discovery or by invention. This is true especially for the scientific and technical knowledge. A researcher typically makes progress by digging deeper within his field of speciality. Sometimes valuable things are also found at the borders of fields or between the fields.

Not every piece of knowledge has an economic value, meaning that we are willing to trade something for it. The value normally depends on the particular situation the user is in, or the task he/she is confronted with. What was of high value this week, may not be the same next week.

The costs associated with a piece of knowledge have very seldom to do with the costs to gain or to produce it. They usually are only the distribution and marketing costs. Considerable parts of knowledge are offered free-of-charge. This may have personal or social reasons. The personal reasons include the pride of the discoverer or owner. The social reasons have to do with the goals of a particular society. Since the beginning of the enlightenment period, knowledge no longer is a privilege of a small portion of society. It is being spread as widely as possible. Education and the possession of knowledge become somewhat like a human right. Society made sure that for certain types of knowledge the access was free, or was kept at a low price. Other information or knowledge was pushed heavily by society (also called propaganda). If knowledge is associated with a price, the upper limit is usually established by the society (i.e. legal constraints) rather than by the forces of the market.

Even if knowledge is free-of-charge, there still may be a considerable effort or a time lag involved before one obtains it. One has to know where to search for it and how to relate it to other knowledge.

3   Electronic markets and the Internet

A market is the (not necessarily geographic) place where the supply and demand of goods and services meet. At the completion of market transactions, goods will be exchanged or services will be negotiated. Each individual transaction can be broken down into four steps or phases [7]: search for information, derivation of decision, negotiation of agreement, implementation. The first two phases can be executed repeatedly.

An electronic market is an information and communication system that supports all or some of the phases mentioned above. It can mainly help in two aspects: facilitate the search for the information and reduce the disadvantages of the geographic or timely distances. If the electronic market is open to the arbitrary participants it can accommodate a larger numbers of participants or increase the degree of specialization of each participant.

The Internet, particularly in form of the World Wide Web (WWW) [1], has all properties of an electronic market and also plays that role. It has achieved that role due to the following circumstances:

- it provides a standardized description of products,

- it allows for a high degree of product details, including pictures and movies,

- it is easy to get access both as a supplier and as a purchaser,

- there exist no timely or geographic restrictions,

- it is possible to conduct within the same medium a large portions of a business transaction,

- there is a considerable public interest in this medium.

There are certain similarities to the mail order business. The advantage is that you do not need to order a catalogue first prior to placing an order. The main problem of the Internet results from its lack of security [2]. The security system of any online market support system has to verify that a certain transaction has really occurred and that it has occurred on behalf of a clearly identified client (unless the electronic money is used). Because of the lack of trust, most financial settlements still occur outside of the network.

Many goods or services are offered free-of-charge in the Internet, because their costs are recovered otherwise, mainly by advertisement. Here the medium is the most important aspect, like in the case of private TV. The goods provided are of secondary nature. Another unique and interesting phenomenon is the fact, that goods and services can be offered close to the

Private notes, collections of measurements
Books, journals, newspapers, reports
Movies, video clips, audio sequences
Equipment plans and instructions
Models of equipment or processes
Patents, invention diclosures
Public statistics, data bases

Figure 1: Expressions of Knowledge.

user without him or her explicitly demanding it. This is called the "push" business.

What counts in case of advertisement (or equally in the case of the paid-for service) is the degree of interactiveness that the user has. He can stay with the provider for hours or can switch off after seconds, leaving no traces or lots of them regarding his identity, preferences and capabilities. This area is largely unexplored, and contains both a potential and a risk.

The electronic market has properties that are both similar and different from the conventional market. First the differences:

- it is easier to provide details or multiple levels of a product description,

- it is more difficult to build up some confidence for a product or for a supplier.

Similarities to the conventional market are:

- customers have to be attracted to the virtual site first, before a product is sold,

- customers should not be brought into a position where they cannot back out from a decision, and

- different modes of payment have to be offered.

These are just examples. We will learn much more about this market as it evolves in the future.

4   The information market in the Internet

In contrast to the physical goods, information products can also be delivered electronically. The same applies to a number of services. Examples of information products are given in Figure 2.

Figure 2 separates the examples in two groups: items that are of general interest to the public at large,

followed by a group of items containing expert knowledge.

The Internet today serves primarily as a source of information about products and services. Millions of suppliers can be accessed, providing tons of descriptive information. This includes private organisations as well as city administrations and universities.

How much business is really made via the Internet is still a matter of debate. The branches that seem to move fastest are those that offer perishable products (flowers, pizzas) or can make business that is otherwise lost (flight tickets) [4]. A very interesting example of the Internet-based businesses are the electronic bookstores. One of them (Amazon books) is managing to offer a selection of 2.5 million titles (impossible to store in a bookstore), at lower prices than other book sellers. Electronic reviews from well-known and unknown commentaries augment the offer. Although clear candidates for the electronic delivery, most books are still print-only and are therefore delivered by airmail or surface mail.

The online information market that largely existed prior to the spread of the Internet, is that of reference and factual databases. About 90 % of that market are covered by economic databases, only 10 % provide technical or scientific material. As far as delivery of digitized maternal is concerned there still is an alternative available in the form of CD-ROMs. They can easily carry data in volumes up to 600 MB. Compared to online data they provide large advantages for the bookstores and publishers. They can be produced, distributed, stored and priced just like books. Also the copyright handling is quite similar. Normally a bookstore is not properly equipped to demonstrate a CD-ROM on a local PC. The buyer has to test the product somewhere else, a major drawback.

From a user's point of view CD-ROMs are only a transport medium. If the information supplied is to be used frequently, it will be loaded on a faster device or

General knowledge:

News, stock prices

Phone books, time tables

Images, films, music

Expert knowledge:

Economic trends, ecological data

Literature references, product suppliers

Molecular structures, reactions

Patents, journals, books

Course material. animations

Figure 2: Information products.

on a local network. This may become quite a problem if a library is offering several hundreds of CD-ROMs. Online distribution of information products of the size required (more than 1 MB) require network capacity that does not exist everywhere. It mainly exists at the universities.

The obvious advantages of either from of electronic supply are:

- Full text searches can be performed

- Color, voice or animation can be added without any extra costs

- Letter size, fonts and overall layout can be easily adapted according to user preferences

- Pictures and references can be shown at the place where they are wanted (rather than several pages later or at the end of a document only)

There are also properties that we are used to, but cannot be transferred to an electronic version easily:

- browsing of the entire document means temporarily transferring the rights for the entire document

- a physical bookmark has to be replaced by a logical substitute, and

- annotations can no longer be made directly in the text.

If we look at the Internet as an information market there are at least the following problems or questions that need to be addressed:

- Since everybody can become her/his own publisher, the quality assurance that is customary for the paper publications may be bypassed.

- Search engines provide large amounts of unstructured data only, unless a good classification scheme is applied by somebody

- Since information sources are maintained at the originator's discretion there is a need for a reliable and long term archiving function.

There are many attempts to deal with these problems. The project described next tries to find some specific answers.

5   The MeDoc approach

5.1   Project participants and goals

Project MeDoc (an abbreviation for Multimedia Electronic Documents) is an attempt to explore the potential of the Internet as an electronic information market. The project is led by an alliance consisting of the German Informatics society (GI), a professional society, the Springer-Verlag, a scientific publisher, and the Fachinformationszentrum (FIZ) Karlsruhe, a database and network services provider. Participants of the project are seven university groups as research partners and another twenty academic institutes and four private companies as pilot users. The goal of the project is to stimulate the transition from printed to

electronic information products in the field of computer science. It is intended to make available full-text and multimedia material for scientists and students at their workplace. The following three subgoals are pursued:

- provide a "critical mass" of high-quality literature in the on-line form,

- evaluate and develop tools and processes supporting an electronic library,

- design an information broker function to guide between heterogeneous Internet resources.

5.2 Online products desired and offered

The pilot users comprise 12 university departments and 8 technical colleges (Fachhochschulen). In the latter case the entire college participates. The four private organizations are all active in the research, training or consulting business. Surveys performed among the pilot users showed that the highest interest was for encyclopaedias (computer science and general), followed by certain core journals in computer science (both national and international ones), and course notes and textbooks for the introductory courses in computer science. Research reports and conference proceedings were considered of lower priority.

It was the user, rather than the supplier, who determined what should be offered, and what should be supplied in which electronic form. In each category, priorities were established with respect to the individual subjects and titles. Very early it became clear that both licensed and unlicensed materials were wanted. As a consequence, a standard licensing agreement was negotiated with the German association of publishers (Boersenverein). This gave the MeDoc alliance a temporary, non-exclusive right to offer certain material in electronic form to the closed community of pilot users.

The MeDoc library currently contains 22 journals and 55 books, originating from 15 different commercial publishers. In four cases, the journals include the current issues. The books comprise some of the best-known German textbooks in introductory computer science, two standard volumes on algorithms and data structures, and several reference books on UNIX, C++, HTML and LaTeX. Of special importance are encyclopaedias of mathematical formulas, some of them illustrated through animations and video clips.

Not for all products which were desired, a license could be obtained nor could all products be offered for which a license is available. Some of the reasons for the first group are:

- The publisher did not have the right for an electronic version himself (typically for some of the books translated from English originals),

- The author did not want the book to be offered electronically without making changes or additions herself/himself

- A new issue was under preparation, or

- The publisher was afraid that the electronic version might impact the sales of the paper version

The second group will be discussed in the next section.

5.3 Conversion to a displayable format

Before telling authors and publishers in what form material should be presented, it was first necessary to develop the concept of an electronic book [3]. In an electronic form, it should be easy to locate things and to review short passages. A displayable book had therefore to be broken down in small segments (chapters, tables, formulas) between which the user should be able to navigate. She should be guided by the table of contents or the index, representing true hypertext links leading to the respective segment. In addition, each book should be related to any other book treating the same or similar topics. This meant full text search over multiple documents.

The desired flexibility could be achieved through either one of two formats: HTML and PDF. HTML allows more choices at runtime and was therefore the preferred format by the users. It has the ability to adapt the layout to a user's needs and to freely include multimedia offerings. PDF, although a proprietary format, keeps the displayed material more closely in synchronisation with the printed version. As a format, it was slightly preferred by the authors and publishers.

Whenever the material existed in the printed form before, a conversion had to take place. It used the author's text format as a starting point, including such formats as Word (multiple versions), LaTeX, Framemaker and others. Only in those cases a conversion was undertaken if the bulk of the material could be translated using a known software tool. Another selection criteria was that either the author (or some other project member) was willing to do the actual work. The result was only accepted if it fulfilled the quality criteria laid out before hand. Several products did not pass this hurdle.

5.4 License administration and usage accounting

Many of the new problems of the electronic market are related to the issue of administration and accounting.

Since little experience is available to draw upon, most approaches are mere trials and exercises. In a project the size of MeDoc, compromises have been made between the desired approach and a workable method. Many of the ground rules we could apply were also predetermined by the rights owner, i.e. the publisher.

As an example, the license for an electronic journal was only given to the subscriber of the paper version. For books, the basic license did not include the facility to print. It was actually a type of subscription, as recommended by O'Reilly [6]. Individual papers of an unsubscribed journal could be requested, but not individual chapters of a book. An important part of each document are the free samples. They entice the potential user to make herself/himself familiar with the type and the quality of the content, before acquiring a license for the actual document.

Figure 3 illustrates the administration and accounting process used in MeDoc. The MeDoc team is the provider. Users can receive a license either as individuals or as a member of a group. In the latter case the administration is done locally. Accounting is done whenever a license is issued. This is not the most flexible form but is quite similar to current practice and easiest to handle. Usage based accounting is much more desirable, but is much more difficult to budget ahead by the user.

To obtain usage related information that is very important, additional statistics had to be collected which are not part of the regular accounting information. They include such items as frequency of use, duration of a session and distribution over time.

5.5   Security Considerations

As far as security is concerned, a balance had to be found between openness for prospects and the interests of the paying user. By making all communication secure would deter the casual visitor from browsing the free and descriptive material.

All information relating to the billing is encrypted, so are any messages relating to the cost producing transactions. If a document is transferred between the server and user it is encrypted as well, so that a third party cannot deduce one's area of interest. The actual billing is done through conventional channels (invoice or paper).

6   Summary and Outlook

The electronic market for information and knowledge products is just developing. What we see today are only its early precursors. To offer products as CD-ROM is a genuine first step, particularly from the suppliers point of view. It avoids some of the problems of a true online market.

The volume of material that appears in form of CD-ROMs is an indication of the potential size of the online information market. It is very obvious that for encyclopaedias and scientific journals the advantages are being exploited. They are caused by technical arguments. Other forms of publications, like research reports and conference proceedings will migrate to an electronic form mainly for the cost reasons. In some cases, particularly for textbooks, the combination of paper and online medium (multimedia supplement) will be the best answer.

Authors are well advised to take the potential of electronic publications into account. Readers have many possibilities to test the new medium. They should do this to help find the optimal solutions. Project MeDoc has taken some steps in this respect, limited to the academic and the computer science field. Its experiences will form the basis of future work in this area.

References

[1] Berners-Lee, T. (October 1996) WWW: Past, Present and Future, IEEE Computer.

[2] Borenstein, N., et. al. (1996) Perils and Pitfalls in Practical Cybercommerce, CACM 39, 6.

[3] Endres, A. (1996) Electronic Books and Journals: Challenges and Opportunities, *Proc. 5th Conference of Greek Academic Libraries*, Thessaloniki.

[4] Komenar, M. (1997) *Electronic Marketing*, Wiley, New York.

[5] Odlyzko, A. M. (1995) Tragic Loss or Good Riddance? The impending demise of traditional scholarly journals Intern. J. Human Computer Studies.

[6] O' Reilly, T. (1996) Publishing Models for Internet Commerce, CACM, 39, 6.

[7] Picot, A., Reichwald, R., Wiegand, R.T. (1996) Die grenzenlose Unternehmung, Gabler, Wiesbaden.
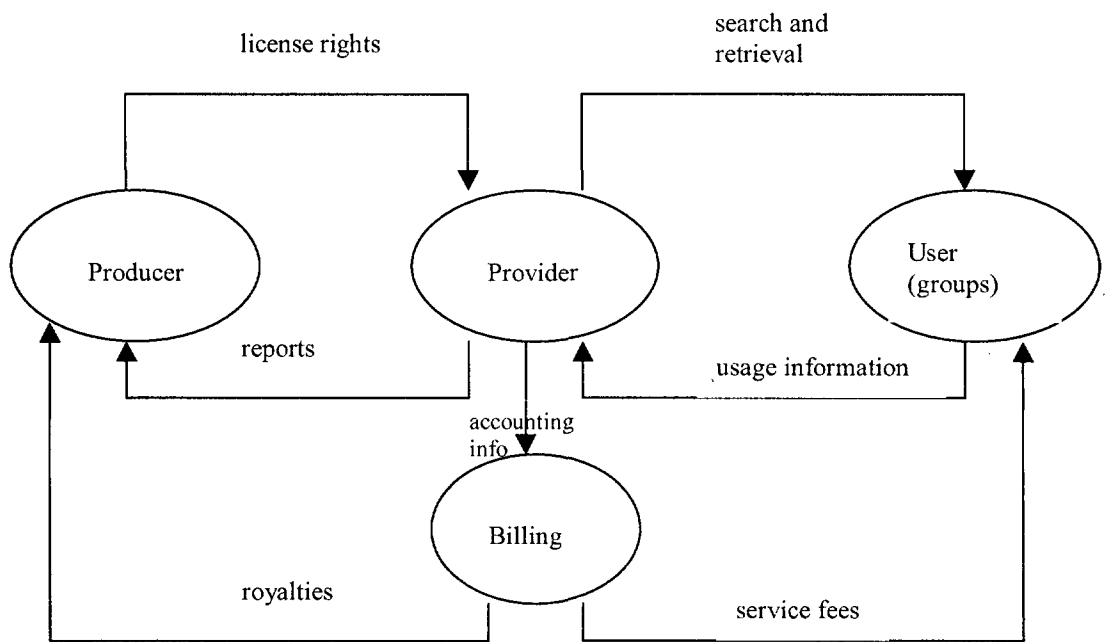
Figure 3: MeDoc administration and accounting process.

Cryptography and Electronic Payment Systems

Janusz Stokłosa
Poznań University of Technology
pl. Skłodowskiej-Curie 5
60-965 Poznań, Poland
E-mail: jsto@agat.sk-kari.put.poznan.pl

The real world payment models and electronic schemes implementing them are presented in this paper. It is outlined that the cryptographic techniques play an important role in the implementations. In the payment protocols appear such cryptographic services as confidentiality, digital signature, integrity, access control and non-repudiation. The authentication of users and the key management are very important as well.

1 Introduction

A payment system carries messages among the subscribers and provides settlement for those messages that constitute funds transfer transactions. Modern wired economy was created when money began turning itself into bits and started flowing around the world through fiber-optics cables and satellite transponders.

Figure 1 presents the contribution of the electronic transactions in the volume and value to the US payments in 1995. A detailed analysis of electronic funds transfer volume and value indicates that they constantly increase. We can expect that the same situation will occur in other countries. These large transactions are guarded by cryptographic mechanisms. The goal of this paper is to give an overview of the cryptographic protection mechanisms and to outline their use in the electronic payment systems.

2 Electronic payment models

There have been a number of proposals for the payment systems. They fall into three broad models: post-paid model, pre-paid model and cash model [6]. The post-paid model is widely implemented in credit/debit mechanisms and includes credit cards, prearranged accounts (for credit), debit cards, checks, pre-paid accounts (for debit). The pre-paid model involves tokens, such as traveler's checks, telephone cards or bank drafts. In the cash model we use cash or gold. Currently proposed schemes of electronic transactions are similar to their real world situations described above as post-paid, pre-paid and cash models. The distinction between on-line and off-line systems comes from the need, or the lack of the need, of a distant computer throughout the transaction. The off-line transaction is more flexible and cheaper.

We would like the electronic money to have the following properties attributed to regular money:

- resistance to forgery,

- privacy,

- digital traveling,

- working off-line,

- transferability to other people,

- divisibility into change.

The digital cash fulfilling these requirements was proposed by Okamoto and Ohta [19]. There exist a number of propositions for electronic payment systems [4-13,18,20]. These schemes use cryptographic mechanisms [23,25] to protect confidentiality and authenticity of transactions. Some of them are briefly presented below.

An iKP ($i = 1, 2, 3$) family of protocols [2] is designed to allow the electronic payment over the Internet with the use of credit cards. The iKP protocol is a public-key based scheme. The parameter i indicates the number of parties (customer, merchant, bank) using cryptographic key. More precisely, cryptographic mechanisms used in iKP protocols are the following: public-key (to protect confidentiality), private key for digital signature, one-way collision-free hash function, and public-key certificate.

Millicent [16] is an on-line payment scheme that uses merchant specific scripts, validity of which is based on hash functions.

Micromint [22] uses collisions in hash functions to create tokens. Since it is hard to come up with collisions in hash functions, only the bank should be able to create tokens.

Mondex [6] is based on a tamper-proof smart card that holds the cash (may be in multiple currencies); software to make and receive payments (or this, authentication techniques should be used). Mondex payments can be made between parties using a hand held "wallet". They can load their wallets remotely from their bank accounts.

3 Cryptographic mechanisms

There is a variety of cryptographic mechanisms. The mechanisms are a basis to special cryptographic services: confidentiality, digital signature, access control, data integrity, authentication exchange, and notarization.

A confidentiality service can be realized by the encryption mechanism. Encryption is a process that conceals the meaning by changing the intelligible messages into unintelligible messages. Decryption is a reverse process to encryption. These two processes form a cryptographic system which is called symmetric (or secret-key) if keys used for encryption and decryption are the same. If the keyes are essentially different, such cryptographic system is said to be asymmetric (or public-key). As far the as symmetric systems are concerned we can use such ciphers as: DES, IDEA, FEAL, Khufu, LOKI, REDOC, Skip-Jack, etc. Among asymmetric ones there are: RSA, El Gamal, and knapsack systems [23,25].

The digital signature lets users verify the source, date and time, and authenticate the contents at the time of the signature. In general, digital signature algorithms are designed as the ones using a public-key cryptographic system. Digital signatures require three operations to be fulfilled: generation of a pair of keys (private and public), signature process with the

private key, and verification process with the public key. The digital signature algorithms can be classified as follows:

- Digital signature algorithms giving message recovery [15],

- Algorithms for digital signature with appendix:

  - Identity-based [4,23],
  - Certificate-based (among them are algorithms based on hash-functions, discrete logarithms and elliptic curves) [1,14,15,17,23,24].

To compute the appendix the use of a hash function is required. We can distinguish the following classes of hash functions:

- Hash functions with cryptographic keys, called also Message Authentication Codes,

- Hash functions without cryptographic keys:

  - Functions using block cipher algorithms [21,25],
  - Dedicated hash functions (e.g., RIPE-MD, SHA) [23].

Authentication exchange can be based on the symmetric and asymmetric ciphers, cryptographic check functions and zero-knowledge techniques [23,26]. In order for a party to authenticate another party they both should use a common set of cryptographic techniques and parameters. The mechanism of notarization protects the information with the help of a third party - a trusted notary. In the process of notarization we can use encryption, digital signature and integrity mechanisms.

The non-repudiation service consists of generation, collection, maintaining, availability, and validation of irrefutable evidence concerning a claimed action or event in order to resolve controversies about occurrence or non-occurrence of the action or event.

As it is mentioned above there are cryptographic mechanisms that need keys. The protection of cryptographic keys is extremely important. Key management deals with the administration and use of the services of generation, registration, certification, installation, distribution, storage, deregistration, revocation, derivation, and destruction of keying material (i.e., keys, parameters, initial values etc.).

	Volume of Transactions	Value of Transactions
Electronic Transactions	$19 Billion	$544 Trillion
Check Transactions	$62 Billion	$73 Trillion
Cash Transactions	$550 Billion	$2.2 Trillion

Figure 1: US payments in 1995, Source: www.nacha.org/uspaymt.htm.

4 Digital signatures

Digital signature is the result of applying the cryptographic transformation to the specific information.

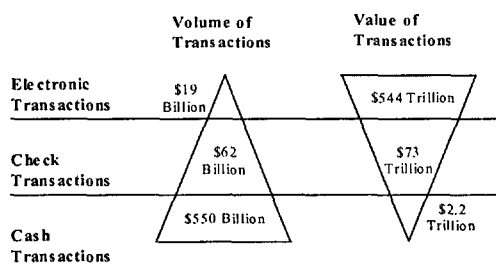For the digital signature, two different keys are generally used: one, called private key, for creating a signature and another, public key, for verifying a signature or for recovery of the message, i.e., returning the message to its original form. It is computationally infeasible [25] to derive the private key from the public key. Computer equipment and software using two such keys is said to create a signature system.

To verify a digital signature, the verifier must obtain a public key and must be sure that the public key corresponds to the signer's private key. To assure that each party is indeed identified with the particular key pair we need a third party trusted by both users of the cryptographic system. The trusted third party is called the certification authority; it is responsible for issuing certificates. To assure the authenticity and inviolability of the certificate, the certification authority signs it digitally.

It is noteworthy that signature systems are requested to provide a relatively short signature for a document of an arbitrary length. It is assumed that the document is first hashed and the signature is produced for its hash code. To make a public key available for use in verification, the certificate may be published in a repository. Repositories are on-line databases of certificates available for retrieval and use in verifying digital signatures. A message bearing a digital signature is verified by the public key listed in a valid certificate. It is as valid, effective, and enforceable as if the message was written on the paper. But the status of legislation is an important problem for effective use of digital signatures in electronic payment transactions. There are certain signature acts established in the world [3]:

- In the USA:

    - Utah Digital Signature Act (1995, amended 1996) can be briefly characterized as follows:
      Policy: facilitate development of Public Key Infrastructure,
      Government agency as top level Certification Authority,
      Comprehensive regulatory scheme (licensing),
      Addresses evidentiary issues,
      Addresses liability issues.

    - California AB 1577 (1995)
      AB 1577 specifies conditions under which a digital signature has the same force and effect as a manual signature. It will be effective as soon as the required regulations are passed, presumably in 1997.

    - Florida Electronic Signature Act of 1996.
    - Illinois Electronic Writing and Signature Act (Draft).

    - Massachusetts Electronic Record and Signature Act (Draft).
    - Other states: Arizona, Connecticut, Delaware, Georgia, Hawaii, Iowa, Kentucky, Louisiana, Michigan, New Mexico, New York, Oregon, Rhode Island, Washington.

- Outside USA:

    - German Draft Digital Signature Law (1997)
      It forms comprehensive regulatory scheme (licensing),
      It does not address evidentiary issues,
      It does not address liability issues.

    - Australian Public Key Authentication Framework.

5 Financial and cryptographic standards

A standard is a documented agreement containing technical specifications or other precise criteria to be used consistently as guidelines, or definitions of characteristics, in order to ensure that materials, products, processes, and services are fit for their purpose. There is no restriction on use of standards. Standardization is the matter of interest of several worldwide and regional organizations. The most important worldwide organizations in the domain of the finances are International Organization for Standardization (ISO, with their Technical Committees JTC 1, TC 68, TC 154), United Nations (EDIFACT), ITU (formerly CCITT), S.W.I.F.T. American institutions have great influence on standards' establishing. Among them are: American National Standards Institute (with its ASC X3, ASC X9 and ASC X12 committees), Federal Reserve, The National Automated Clearing House Association, American Bar Association. Standards dealing with cryptographic techniques are developed since 1977, beginning with the US Data Encryption Standard (DES). From the seventies the standardization in cryptography has developed intensively. Here, the most important institution is JTC 1 of ISO and IEC (International Electrotechnical Commission). The scope of the subcommittee ISO/IEC JTC 1 SC 27 includes:

- Identification of generic requirements for information technology system security services.

- Development of security techniques and mechanisms, including registration procedures and relationships of security components.

- Development of security guidelines, e.g., risk analysis, interpretative documents.

– Development of management support documentation and standards, e.g., security evaluation criteria.

6 Final remarks

Payment mechanisms and services are used, and will be used intensively in payment systems implemented in computer networks. To ensure the security of transactions, cryptographic techniques such as confidentiality, digital signature, integrity, non-repudiation are used. The authentication of users and the key management are very important as well. The role of standardizing organizations is significant.

Cryptography offers mechanisms for implementations of concrete payment protocols. These mechanisms employ a variety of mathematical and physical tools to protect financial messages sent via computer networks. However, it is important to use particular mechanisms to manage encrypted information properly.

References

[1] Agnew G.B., Mullin R.C., Vanstone S.A. (1990) Improved digital signature scheme based on discrete exponentiation, Electronics Letters, 26, 14, p. 1024-1025.

[2] Bellare N., et al. (1995) *iKP - a family of secure electronic payment protocols, Proceedings of Usenix Electronic Commerce Workshop*, p. 1-20.

[3] Biddle B. (1997) Legal issues raised by digital signature and PKI: an overview, *The 2nd Meeting of the Electronic Payments Forum*, La Jolla, CA.

[4] Boly J.-P., et al. (1994) The ESPRIT project CAFE - High security digital payment systems; (ed) D. Gollmann, *Computer Security - ESORICS 94*, LNCS 875, Springer-Verlag, Berlin, p. 217-230.

[5] Brands S. (1994) Untraceable off-line cash in wallet with observers, (ed) D. R. Stinson, *Advances in Cryptology - CRYPTO'93*, LNCS 773, Springer-Verlag, Berlin, p. 302-318.

[6] Buck S.P. (1995) Effecting transactions on the superhighway or Making payments on the Internet, Hyperion Systems Ltd., http://www.hyperion.co.uk/pub/library/-HTMLLibrary/SPBPayments/Payments

[7] Chaum D. (1985) Security without identification: Transaction systems to make Big Brother obsolete, Communications of the ACM, 28, 10, p. 1030-1044.

[8] Chaum D., Fiat A., Naor M. (1990) Untraceable electronic cash, (ed.) S. Goldwasser, *Advances in Cryptology - CRYPTO'88*, LNCS 403, Springer-Verlag, Berlin, p. 319-327.

[9] Damgard I. B. (1990) Payment systems and credential mechanisms with provable security against abuse by individuals, (ed.) S. Goldwasser, *Advances in Cryptology - CRYPTO'88*, LNCS 403, Springer-Verlag, Berlin, p. 328-335.

[10] Even S., Goldreich O., Yacobi Y. (1994) Electronic wallet, (ed) D. Chaum, *Advances in Cryptology: Proceedings of Crypto 83*, Plenum, New York, p. 383-386.

[11] Ferguson A. (1994) Extensions of single-term coins, (ed) D. R. Stinson, *Advances in Cryptology - CRYPTO'93*, LNCS 773, Springer-Verlag, Berlin, p. 292-301.

[12] Fuchsberger A., et al. (1996) Public-key cryptography on smart cards; (eds.) E. Dawson, J. Golić, *Cryptography: Policy and Algorithms*, LNCS 1029, Springer-Verlag, Berlin, p. 250-269.

[13] Hirschfeld R. (1993) Making electronic refunds safer, (ed.) E. F. Brickell, *Advances in Cryptology - CRYPTO'92*, LNCS 740, Springer-Verlag, Berlin, p.106-112.

[14] IEEE P1363 Working Draft, August 18, 1997.

[15] ISO/IEC 9796, Information technology - Security techniques - Digital signature scheme giving message recovery, 1991.

[16] Manasse M.S. (1995) Millicent (electronic microcommerce), http://www.research.digital.com/SRC/personal/-Marc_Manasse/incommon/ucom.html

[17] Nyberg K., Rueppel R. A. (1993) A new signature scheme based on the DSA giving message recovery, *1st ACM Conference on Computer and Communications Security*, Fairfax, Virginia, 1993.

[18] Okamoto T. (1995) An efficient divisible electronic cash scheme, (ed.) D. Coppersmith, *Advances in Cryptology - CRYPTO'95*, LNCS 963, Springer-Verlag, Berlin, p. 438-451.

[19] Okamoto T., Ohta K. (1992) Universal electronic cash, (ed.) J. Feingenbaum, *Advances in Cryptology - CRYPTO'91*, LNCS 576, Springer-Verlag, Berlin, p. 324-337.

[20] PfitzmannB., Waidner M. (1992) How to break and repair a "provably secure" untraceable payment system, (ed.) J. Feingenbaum, *Advances in Cryptology - CRYPTO'91*, LNCS 576, Springer-Verlag, Berlin, p. 338-350.

[21] Pieprzyk J., Sadeghiyan B. (1993) *Design of Hashing Algorithms*, LNCS 756, Springer-Verlag, Berlin.

[22] Rivest R.L., Shamir A. Payword and Micromint: Two simple micropayments (unpublished).

[23] Schneier B. (1994) *Applied Cryptography*, Wiley, New York.

[24] Schnorr C.P. (1990) Efficient identification and signatures for smart cards, (ed.) G. Brassard, *Advances in Cryptology - CRYPTO'89*, LNCS 435, Springer-Verlag, New York, p. 239-252

[25] Stokłosa J. (1994) *Cryptographic Algorithms* (in Polish), OWN PAN, Poznań.

[26] Wayner P. (1996) *Digital Cash: Commerce on the Net*, AP Professional, Boston.

Database Support for Intranet Based Business Process Re-Engineering

Wojciech Cellary, Krzysztof Walczak and Waldemar Wieczerzycki
Department of Information Technology, University of Economics
60-854 Poznań, Poland
E-mail: `cellary@kti.ae.poznan.pl`

In the paper a new approach to the organization and management of business processes by enterprises which use advanced information technologies is presented. The proposed data model allows flexible modeling of processes which can evolve over time, thus precisely reflecting dynamic changes of the enterprise management procedures. A single process model may be available in different variants. These variants reflect alternatives in the management procedures which relate to exceptions, different triggering events, different time periods, resource limitations, etc. The proposed approach is based on both the database technology and the intranet technology, thus enabling practically unlimited flow of information between all employees of an enterprise, coordination of their collaborative work, decision support and wide access to the legacy systems.

1   Introduction

Efficient business process management and information flow are the crucial factors which directly influence achievements of all enterprises. The importance of these factors becomes more visible in case of big and geographically distributed enterprises which carry out versatile activities. Efficiency of both business process management and the information flow strictly depends on the technology of information exchange being used by an enterprise. Up to date, the flow of printed paper materials (e.g. letters, faxes, reports, financial records, leaflets, technical documentation) complemented by the direct personal meetings and phone calls has been the most common technology of information exchange. Nowadays one can observe the emerging modern media and communications technologies which enable to organize the information exchange in a different way. We mean here electronic multimedia documents which replace the paper documents and video-conferences which replace direct personal meetings, no matter what is the geographical distance between the participants. In comparison to a paper document, an electronic one can be easily modified and versioned, it can be easily transmitted to a recipient or a group of recipients of an arbitrary size, an it can be automatically processed to acquire some knowledge. Video-conferencing eliminates costs of traveling and accommodations and saves time. In such situation a natural question arises: can the new communications technologies improve the business processes of an enterprise and its management?

Most of problems faced by the enterprises concern internal business processes that are neither well defined nor particularly efficient. Organizational procedures as well as service to customers need to be improved, while difficulties in introducing changes within the enterprise should be minimized. A workflow management system in which the concepts of business process modeling and reengineering are applied, based on a computer network, is a potential solution to these problems [2,3,4,8,9]. Up to date, the main obstacle of the development and wide use of such systems was the lack of a technology providing a possibility of their deep and flexible adaptation to the enterprise needs. Hopefully, nowadays this obstacle has disappeared, due to the intranets. Intranet is an internal enterprise Internet. It is a use of Internet technology to organize the work of an enterprise. Intranet works on a private enterprise network, it is appropriately secured, and usually disconnected or connected in a limited way to Internet [7]. There are multiple advantages for enterprises from the use of the intranet. Some of them come from the technology itself, but majority are a result of changes in the way, the enterprises do their work. The two main advantages of intranet are [17, 18]:

- faster and easier access to more accurate company information

- faster and better communication among employees.

The above advantages imply that the intranet is cur-

rently the main focus research of many information scientists, developers, and software production companies. These advantages, however, are still potential, since intranet technology is very young and it lacks applications reflecting the specificity of the particular enterprises. Thus, there is an open, wide area for software designers who can improve business processes of enterprises using this very powerful technology.

The main goal of this paper is to show how business processes can be organized and managed by enterprises using advanced information technologies. We assume flexible process models which can evolve over time, thus precisely reflecting dynamic changes of the enterprise management procedures. The evolution of a process model can be multi-thread which means that we allow a single process model to have many variants. These variants reflect alternatives in the management procedures which relate to exceptions, different triggering events, different time periods, resource limitations, etc.

The main technical concept of the proposed approach is to merge the database technology with the intranet technology, thus enabling practically unlimited flow of information between all employees of an enterprise, coordination of their collaborative work, decision support and - what is even more important - wide access to the legacy systems. The intranet makes it possible to preserve and access all information resources of the enterprise, no matter how old they are. The same concerns the hardware and software systems which can be integrated with new technologies in a way guaranteeing their full accessibility.

The paper is structured as follows. In Section 2 we present a particular approach to modeling and executing the business processes in a computer-based information system. In Section 3 we show how a multi-version database can be applied to support the model proposed in Section 2. In Section 4 we explain how the proposed approach can be practically implemented on a computer platform equipped with the intranet. Section 5 contains concluding remarks and some directions of future work.

2   Description and Execution Model of Business Processes

In a real life an enterprise is involved in many business processes which aim to achieve particular goals. One may perceive these processes at two different levels: conceptual and execution. At the *conceptual level* business processes exist in a generic and abstract form as a universal description (printed, oral or electronic), which specifies how the processes should be executed in a case of a real need. This description contains the structure and behavior of all objects that can potentially be used during business process execution,

or produced as a result of its execution (e.g. people, machines, tools, forms and documents). It also contains the exact specification of all activities embedded in a business process, their scheduling and activation rules, and their inputs and outputs. At the *execution level*, instead of a pure description, every business process becomes a real work performed by the enterprise employees, according to this description. In general, a single business process from the conceptual level corresponds to many business processes at the execution level. For example, a *client acquisition* business process can be seen at the conceptual level as a generic description of how to attract a new client to the enterprise. At the execution level this single conceptual business process becomes a set of executional processes, since actually the enterprise is doing its best to attract *Mr. Smith*, *Mr. Jones* and *Mrs. Brown*.

In our approach, at the conceptual level an enterprise is represented in the information system by so called system background and a set of process models [15]. The *system background* models an enterprise without the business processes. It contains all environments which may be used to create the process models, which in turn are composed of objects, actors, available resources, standard roles etc. A *process model* is a precise description of a corresponding real-life business process. Since the enterprise represented in the system evolves over time to maximize its efficiency, profit, quality of services, etc., process models also evolve, thus ensuring consistency between the real-life processes in the enterprise and the process models in the system. This evolution is achieved step-by-step through the creation of improved versions of the process models. New versions of process models, however, do not just replace older versions - they are kept in parallel to them. Preserving old process model versions (called *revisions*) makes it possible to perform roll-backs in process model evolution, for example due to the unsatisfactory results obtained. On the other hand, sometimes there is an evident need to return to the previous versions of process models, due to the change of requirements or other factors which have a cyclic nature, e.g., the most recent process model version is adequate for every month except of December, when a former version is more adequate.

There is also another aspect of process model versioning. Sometimes, a single process may need to be performed in a slightly different way depending on triggering events, particular conditions, temporary results obtained, etc. In this case, instead of a single, the most recent version of a process model, we deal with a number of its *variants*. Contrarily to revisions which reflect the progressive nature of processes, variants reflect their alternative nature. The following three events related to process models can occur in the system:

– process initialization,

- process definition,

- process instantiation.

Two first events relate to the conceptual level, while the third one to the execution level.

Process initialization consists in the creation of a new process model, which initially is available in a single version. This process model version contains logical copies of some objects contained in the system background, i.e. the process model is directly derived from the background. All process models are initialized (derived) from the same background by selecting subsets of its objects. These subsets need not be disjoint, i.e. some objects stored in the background can be shared by the different process models. Update of the background objects is allowed. Modifications are automatically propagated to those process models which contain the objects affected. On the contrary, newly created background objects are not included in the process models formerly initialized. They can be used in subsequent initializations of new processes. Updates in the process models are local, which means that they do not influence the background and other processes. The *process definition* consists in refining a process model. It is done by the derivation of a new process model version, which next can be freely modified. In particular, it can be extended by a specification of all the activities that are comprised by the process, which normally are not included in the background. It can be also extended by the description of objects local to the process model version. Finally, semantic relationships among the objects mentioned above can be introduced, which order activities, assign them particular roles, point objects that are exchanged between activities or are used internally by an activity, bind external and internal events to actions which they trigger, etc. The process definition can be performed in many steps, which means that the total number of process model versions is practically unlimited. The concepts introduced so far are illustrated in Figure 1. In the system two real-life business processes are modeled: P_1 and P_2, which were initialized by getting two subsets of objects included in the background. The process model P_1 is composed of five versions, while the process model P_2 is composed of six versions. Model versions of P_1 form a chain with the most recent version being derived subsequently from its four revisions. Model versions of P_2 form a tree with two branches which are variants.

Let us remind that the concepts introduced so far relate to the conceptual level, since they reflect static features of a real-life enterprise. Now, we will focus on the concepts related to the execution level which reflect dynamic features of the enterprise. In our approach, an enterprise considered at the execution level is represented in the information system by so called process instances. A *process instance*, or shortly -
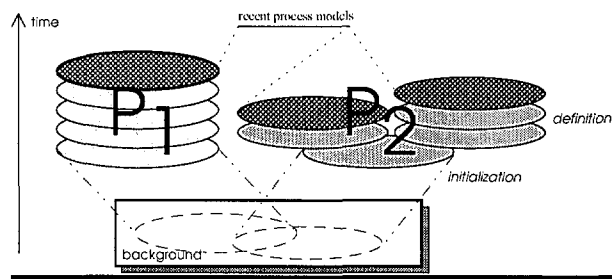

Figure 1: Business process initialization and definition.

a *process*, represents a single execution of a corresponding business process, according to a selected process model version. In general, every business process (e.g. advertising) can be performed in parallel as many times as required (e.g. a new car can be advertised in parallel with a new motor bike). Thus, a single process model version can have an arbitrary number of instances, which can run in the system simultaneously and asynchronously.

The creation of process instances is not related to the two events which can occur in the system, introduced so far (i.e. process initialization and process definition). It is triggered by the third event, namely by the process instantiation. Briefly saying, the *process instantiation* starts the execution of activities embedded in the process, according to the patterns included in the corresponding process model version. In general, the execution of one process can be performed in parallel to the execution, initialization, and definition of other processes. The main aim of the proposed approach is to avoid conflicts between these operations. The majority of conflicts arise when different processes try to access the same object. The most straightforward solution to avoid those conflicts is to isolate processes addressing them to different, logically independent workspaces. It is relatively easy in the case of instances of different process models, while more difficult in the case of the processes being instances of the same process model version. To avoid conflicts between instances of the same process model version, a special new version of this process model is automatically derived whenever a process is instantiated. This new version is temporary and becomes an exclusive scope in which the newly instantiated process is being executed. Initially, it comprises the logical copies of all objects included in the parent version, i.e. it contains all patterns included in the selected process model version. Afterwards, it can evolve due to changes implied by process execution. Process model versions representing different process instances are logically independent, even if they correspond to the same process model version. In the latter case, instances of the process are direct children of the same node which models the respective process itself.
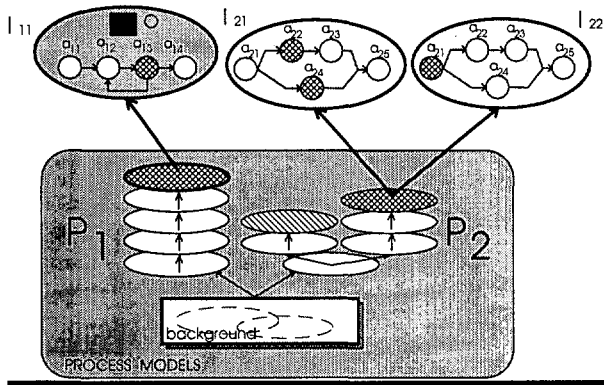
Figure 2: Business process instantiation.

Every process can be composed of an arbitrary number of elementary activities which are partially ordered. It is also assumed that, in general, processes are interactive and require intensive information exchange with users, who are modeled in the system as actors. One may distinguish three types of processes executed in the system. The first type contains read-only processes that read objects stored in the system, but do not update them and do not create new objects. Of course, processes of this type can produce new information, e.g. reports, letters, management directives, etc., which is displayed or printed to the users. Processes of the second type can create new persistent objects which are local to them, i.e. which exist only during the process execution and play the role of outputs/inputs exchanged between the activities in the scope of the same process. Processes of the third, the most general type can both update the objects existing in the system and create new persistent objects that are retained in the system after process execution. The above concepts are illustrated in Figure 2. The most recent model version of process P_1 has been instantiated only once, while the model version of process P_2 has been instantiated twice. The single instance I_{11} of P_1 is a process of type three, since it has produced a new persistent object O. It is composed of four activities: a_{11}, a_{12}, a_{13}, and a_{14} with a_{13} being currently executed. Both instances of P_2: I_{21} and I_{22}, which were derived from the same version of the process model, are instances of type two (i.e. they do not produce persistent objects). Since they are logically independent, they are executed in mutual isolation. Process instances I_{21} and I_{22} have the same schedule of activities. They are composed of five activities: a_{21}, a_{22}, a_{23}, a_{24} and a_{25}, which are partially ordered. I_{21} and I_{22} are executed asynchronously. In case of instance I_{21}, its two activities a_{22} and a_{24} are currently executed in parallel. In case of instance I_{22}, activity a_{21} is currently executed.

The following question arises: what to do with the temporary versions representing processes after they are finished? In the case of both processes of the first and the second type (as distinguished above), those versions can be simply removed from the system. In the case of the process of type three one must decide what to do with objects created and modified by this process. In general there are two possibilities:

– the temporary version is promoted to a persistent one, i.e., it becomes a new version of the process model, uniquely identified and visible to the users,

– modifications made in the temporary version are moved to the parent version, by *merge* or *redo* operations [16], and child version is deleted.

If temporary versions are systematically promoted to the persistent ones then the number of versions of the same process model increases very rapidly. Since there is definitely a difference between a version which represents a modified process model and a version which has been created due to modifications of system objects via such a process, two levels of version addressing have to be available to the user. The first level concerns the versions corresponding to process modeling operations (i.e. process initialization and definition), while the second one concerns the versions created as instances of the same process, i.e. versions sharing the same process model and varying only by objects being the deliverables of the corresponding process.

3   Multiversion Database Support

The description and execution model of business processes presented in Section 2 needs a support of a database management system, since such a system provides object persistency and concurrency control. Recently available object-oriented database management systems provide also object versioning mechanisms which can be exploited for representing versions of the process models and instances. Those mechanisms, however, are not the same in different database management systems, which means that some of them are more appropriate for our approach than others. In this section we briefly classify approaches to version modeling and management in object-oriented databases. Next, we select the approach which, to our opinion, is the most relevant, and finally we present how the business process modeling concepts can be mapped into the concepts of this approach.

Intuitively, a multiversion database is a persistent object storing system in which every object can be potentially represented by an arbitrary number of its versions. In comparison to a conventional (i.e. monoversion) database, a multiversion database adds

a new granule of data storage, namely an object version, which is somehow embedded in a corresponding multiversion object. Moreover, a new semantic relationship between versions of the same object is kept in the multiversion database, respectively for every database object. There is also another important difference between monoversion and multiversion databases which relates to the problem of finding subsets of database objects that "go together". As it is well known, whenever a transaction is addressed to a monoversion database, the entire database is assumed to be in a consistent state, i.e., in a state that corresponds to one of valid states of the real- world being modeled. Contrarily to monoversion database, a multiversion database is in general not consistent, according to the classical definition of consistency. More precisely, considered as a whole, it does not correspond to any valid state of the real-world. That is why, one must distinguish particular subsets of object versions that are mutually consistent in order to address to them users' transactions, which are assumed to preserve the consistency of those subsets after their commitment. As related to this issue, the existing version models can be divided into three families, according to the versioning granularity.

In the first family [1,11,14], a multiversion object is a versioning granule. It implies that to associate consistent object versions, it is necessary to explicitly link versions: object versions can be referenced by other versions to build the complex object versions. This link is used once as a reference of an entity to another, and once to associate two mutually consistent entity versions. This duality generates problems to manage consistent object version sets: creation of object versions implies creation of a new complex object version, mainly achieved manually or by automatic generation of many, often useless, complex object versions. Thus, those works are limited both by the number of complex object versions managed simultaneously and by the functionalities to manipulate them.

In the second family of version models [10,6], generally devoted to the design applications, a subset of objects, called generic configuration, is a unit of both the versioning and the consistency. This implies that a new object version may be created only inside a particular configuration (i.e., a version of a generic configuration), thus avoiding the problem of explicit linking of mutually consistent object versions. However, this approach is restrictive, since it usually imposes the use of a top-down design. Moreover, sharing objects between different generic configurations raise difficulties similar to those appearing in the first family.

Finally, in the third family, the complete "monoversion" database is a versioning unit. Units of consistency across versions are thus composed of versions of all the objects in the database, one version per object. This approach, in some sense, is used for example in temporal databases [12,13] which represent different dimensions of time (e.g. validity time, transaction time etc.). If only one dimension of time is represented (for instance the validity time) the database stores as many states of the real world as there are different validity states present in the database. However, the total order of versions imposed by the semantics of time flow induces a strong limitation on applying this approach to other domains, where the versions are generally organized in a tree or a DAG, instead of a sequence. On the contrary, in the database version model [5], the organization of versions coming from the derivation is not limited to a sequential order. This model is very straightforward and natural, since it reflects both progressive and alternative nature of real-world processes being modeled in the database. These two important advantages have convinced us to use the database version model in the further discussion.

As we have mentioned, a conventional monoversion database represents one state of a modeled part of the world. According to the database version approach, a *multiversion database* represents simultaneously several states of the modeled part of the world. Each state is represented by a *database version*, as shown in Figure 3. Each database version, denoted dv, has an identifier. It contains a version, called the *logical object version*, of each object stored in the database. Objects are *multiversion*, i.e., they are composed of several logical object versions. A logical object version is similar to an object in a monoversion database: it has an identity and a value. The identifier of a logical version of a multiversion object, contained in a database version, is a couple $<multiversion\ object\ identifier,\ database\ version\ identifier>$. To express the fact that conceptually an object does not exist in a particular database version, the value of its corresponding logical version is set to $+$, meaning "doesn't exist". Different logical versions of an object can have the same value. In such case this value is stored only once in a *physical object version* and it is shared by several logical object versions. A physical object version has a unique identifier. The database management system controls the association of logical and physical object versions. Given a set of database versions, the creation of a new database version dv_n is achieved by logically copying one of the existing database versions. This process is called derivation. After derivation, users may update logical object versions contained in the new database version at any time. A modification of a logical object version is followed by updating its corresponding physical object version. If a physical object version is associated exclusively with the modified logical object version, it is simply updated, otherwise, i.e., when it is shared by several logical object versions, a new one is created where the modifications are stored.
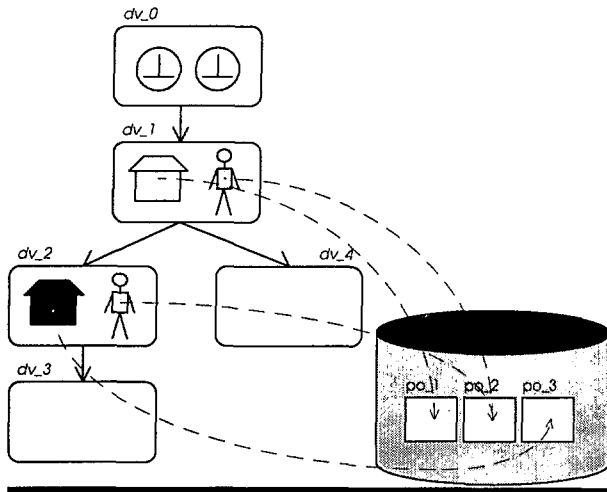
Figure 3: A multiversion database.

In a conventional monoversion database an object can be referenced by an arbitrary number of objects and can in turn reference an arbitrary number of other objects. In the database version approach dynamic binding is used to bind a versioned object with another versioned object. A reference to an object version, stored in a physical object version, contains only the identifier of a referenced multiversion object, while its version is left unspecified. The database management system selects the default version identifier at the logical level. The default version identifier is the database version identifier which is common to all logical object versions contained in the database version. This way of referencing is a key issue of the database version approach, because it makes it possible to update database versions independently of each other. In Figure 3, the multiversion database is composed of five database versions. Database version dv_1 is derived from dv_0. Database versions dv_2 and dv_4 are both derived from the database version dv_1, as two alternatives. Database version dv_3 is derived from database version dv_2. The multiversion database contains two multiversion objects: a *house* and a *person*. The color of the *house* is different in dv_1 and dv_2, so, its two logical versions are bound with two different physical house versions po_1 and po_3, respectively. The *person* remains the same in both database versions dv_1 and dv_2, so, its logical versions share the same physical person version po_2.

Assuming a multiversion database managed according to the database version approach as a basic platform for business process modeling and execution, we can now easily map the concepts introduced in Section 2 into the database concepts. Since database versions are logically independent and they are the units of database consistency, the first basic idea is to represent all process model versions by independent database versions. The background modeling the

enterprise without activities (cf. Section 2) is represented by a single database version being the only child of the root database version, i.e. being the only node of the 1st level of database version derivation tree (cf. Figure 4).

Every business process is represented in the database by a separate subtree of database versions. The root of every subtree is a direct child of the database version representing the background, i.e., it is a node of the 2nd level of database version derivation tree. Each database version included in the subtree representing a single business process corresponds to a single version of the process model. As a consequence, the process initialization consists in deriving a new database version of the 2nd level. Since at the very beginning this database version is a logical copy of its parent, some of its objects (e.g. useless objects) must be then updated by the + value, while other preserved without changes.

Similarly to the process initialization, the process definition consists of deriving the new database versions representing the new versions of the process model in the scope of the respective database version subtree. At this stage new objects are introduced to the database versions which are local to the corresponding business process. Also objects initially copied from the background can be modified in order to reflect the specificity of the business process. Finally, process instances are also represented by the database versions. Contrarily to the database versions modeling business processes, these database versions are temporary and can be kept on a server, rather than in the database itself. All instances of the same process versions are direct children of the database version storing the process model (cf. Figure 4). They physically share all objects describing the process with their parent, and can introduce (or modify) objects related to a respective process execution. Thus, the process instantiation can be, again, perceived as the database version derivation, with the only difference that a newly derived database version is not persistent, and can not be further used as a parent node for database version derivation.

All the mappings proposed above are illustrated in Figure 4. The multiversion database is composed of 10 database versions. 1st level database version identified by dv_1 represents the background. Two subtrees of the derivation tree, rooted by two 2nd level database versions: dv_2 and dv_3, represent two business processes: $proc\_1$ and $proc\_2$, respectively. The model of business process $proc\_1$ is available in three different versions, while the model of $proc\_2$ in two versions only. $proc\_1$ is well defined, however not executed at this time. $proc\_2$ is not only well defined, but also currently executed. Its three asynchronously created instances are represented by three temporary database versions, being the following 4th level nodes:
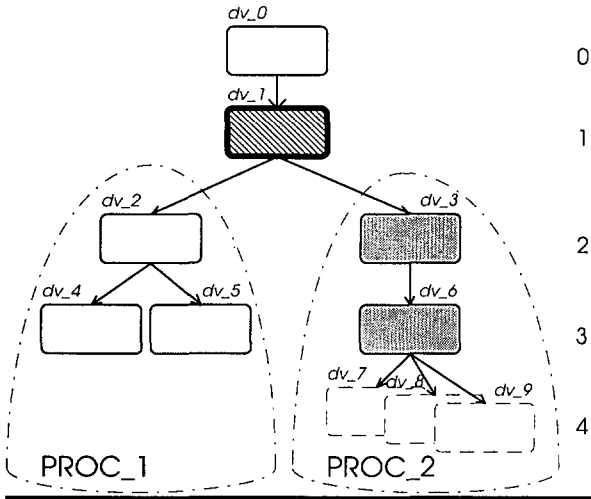
Figure 4: Process modeling and execution in the multiversion database.

Figure 5: Temporary database versions.

dv_7, *dv_8* and *dv_9*.

Since the temporary database versions are not stored in the database, at the end of execution of the corresponding process one has to decide whether they are simply deleted or additional actions are undertaken by the management system. In case of a process which does not produce persistent objects as the results of its execution (i.e. objects that should be kept in the database), the respective temporary database version is just deleted from the system. In case of a process which modifies or/and introduces objects which refine the process model, a new process model version is created. That is why at the end of the process execution the corresponding temporary database version is promoted to a persistent database version, and its relationship with the parent database version is preserved. Finally, in case of a process which does not change its definition, but creates new objects being the outputs of process execution, the management system can automatically *redo* all its operations in the parent database version, thus introducing these objects to a persistent database version, and next delete its respective child. This *redo* operation is based on the *log file* and is performed in a serial way. It means that in case of more than one process of this type, no conflicts arise in the parent database version during the redo operation.

To illustrate the above mechanisms let us consider the *client acquisition* process of a particular enterprise which has four instances, as shown in Figure 5. Instance *i1* fails, i.e. no new client is added to the database. Thus, the respective database version is simply deleted from the system. Instances: *i2* and *i3* succeed, since they have attracted two new clients to the enterprise. In such situation, transactions committed in the corresponding database versions are serially *redone* by the management system in the parent database version and the two children are deleted. Fi-
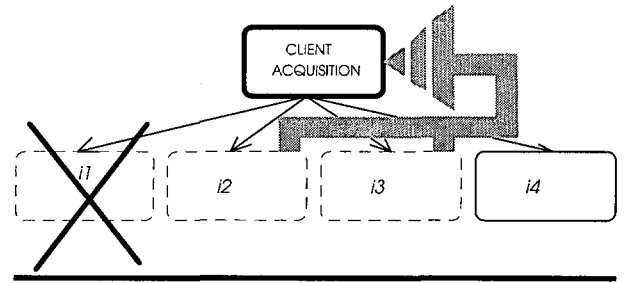
nally, instance *i4* improves the strategy of the client acquisition process, in order to avoid future failures, by adding new information resources and upgrading the technologies being used. As a consequence, a new version of the process model is created, which means that the corresponding database version has to be promoted to the persistent database version.

4 Combining business process modeling technique with intranet technology

In the previous section we have discussed how a multiversion database management system can be used for representation of process models. A sole database system, however, may not be sufficient when a real enterprise implementation comes into consideration. There are few issues which definitely have to be addressed:

- distribution - enterprise employees usually work on different computers. Sometimes all these computers are interconnected and form a local area network (LAN). There is a great number of enterprises, however, which have more complex, distributed structure with many divisions in distant places, usually not interconnected by direct LAN links,

- diversity - currently in many companies there are multiple coexisting different computer systems. Thus, it is crucial for the implementation policy to be able to incorporate and reuse to the maximum these legacy hardware and software elements,

- multimedia - computer users became very demanding. Nowadays multimedia information is an important element of the computer systems and it is difficult to imagine a new and successful primary enterprise system implementation without support for multimedia data.

There is currently only one networking technology which can meet these requirements - intranet. Intranet is a network environment based on the Internet standards, which is usually isolated, or connected in a lim-

ited way to the Internet at large. This environment is owned by the enterprise and usually not accessible from the outside. It consists of a computer network and a set of software tools which enable its efficient employment [17, 18, 19].

Recently, intranets gain more and more acceptance as an integrated solution for the enterprise networking. There are multiple reasons for this tendency. Proprietary groupware systems as opposed to intranets are often limited in their capabilities and usually difficult to improve. It is difficult to follow the rising user requirements with such proprietary solutions. As a rule, these systems are platform dependent. Changing the software layer requires major investments in the computer hardware, and usually results in a period of instability and reduced effectiveness of the enterprise.

Over the years of development and extensive utilization, Internet proved to be a convenient and robust environment. Primarily it was used only for the information distribution, but recently with the appearance of Web programming techniques it can be very effectively used also as an application platform. To further advantages of intranet technology the following can be included:

- intranets are scalable,

- intranets are simple,

- intranets are based on standards.

Scalability, in this context, means a possibility to increase the amount of data and services available within the system with little or no effort needed to redesign its structure. Scalability is a result of highly distributed architecture of the network and use of software tools which were designed to work with limited bandwidth and very high load conditions which are common on the Internet. Simplicity was one of the main reasons for Internet success. The browser software itself provides only a minimal user interface leaving the actual presentations to be built by the content developers. Important element of the Web technology significantly simplifying its usage are hyperlinks. Hyperlinks allow users to easily navigate and find nformation by simply clicking a word or graphics. Sending an e-mail or downloading a file to be displayed is no more complex. Web technology is based on open standards. Client software from any vendor can use the information served by a server from other software manufacturers. Web tools are available for all popular hardware platforms and operating systems and usually can be connected to the legacy database or other back-end systems of the enterprise.

The basic structure of intranet is presented in Figure 6. In general, it consists of one or few intranet servers, optional Internet connection through a special proxy server (for security reasons), and multiple clients being work- and-access platforms for users. The
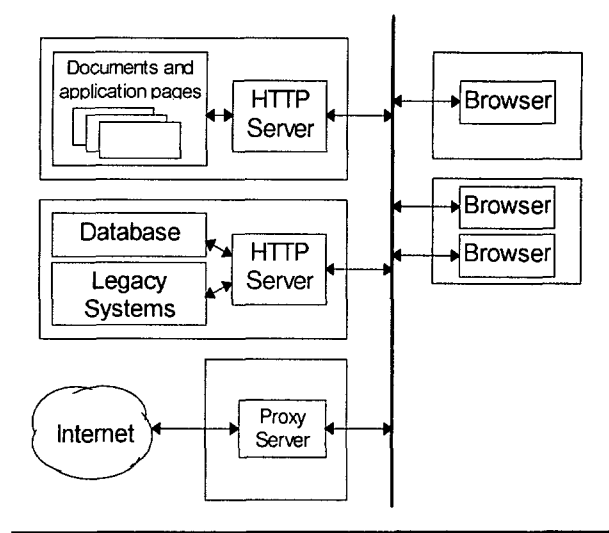

Figure 6: Basic structure of an intranet.

functionality provided by the servers and clients is still becoming richer and richer. Functions such as browsing and editing documents, receiving and sending e-mail messages, downloading files, and access to newsgroup lists is currently rather common for all major Web browsers. Moreover, contemporary browsers offer capabilities far beyond simple document manipulation. They can be used as an execution platform for application modules (e.g., Java applets) contained in the Web pages downloaded from the server. This feature combined with dynamic creation of Web pages on the server site gives a very powerful mechanism of controlling contents, configurations, or versions of applications delivered to the user. On the server site, sending documents on browser requests is the most basic function. More advanced ones include access to databases and execution of programs which then allow to access legacy software systems of the enterprise (e.g., through CGI).

The great flexibility of Intranet comes partially from its structure:

- clients and servers can work on various platforms - they are independent and communicate only by commonly accepted standard protocols,

- each user connects to the system by the use of a universal browser, which is a platform providing all necessary system functions in a unified way on different operating systems,

- centralization of all kinds of data and data access requests in servers makes it possible to control the document flow,

- distribution and independence of servers gives simple way of scaling the system in case of a great growth of data and traffic volume.
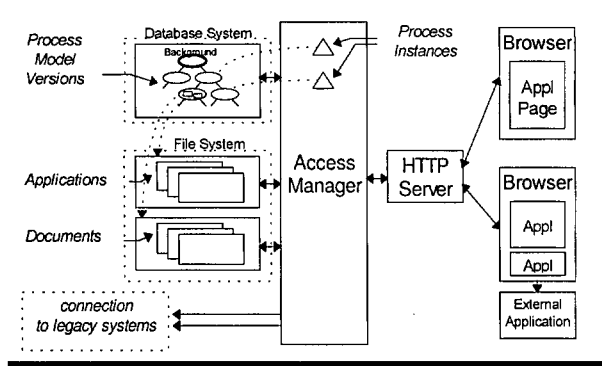
Figure 7: Structure of an Intranet-based system for business process reengineering.

Figure 8: Distribution of process model database.

Structure of an intranet-based system for the use in business process reengineering is presented in Figure 7. Browsers serve as a basic document manipulation and application execution platform for the system users. Functions which are not provided by the application pages running inside the browser are accomplished by the automatically launched external applications. Browsers connect to an HTTP server which is used as the main document and application delivery mechanism. HTTP server has access through the Access Manager to all documents and application pages. It is also connected to the database where all the process models are stored. Access Manager can also communicate with the legacy software systems. Instances of business processes are created inside the Access Manager module. They can affect the way the Access Manager chooses appropriate versions of database or files.

In addition to the general advantages of intranet, the system configuration presented in Figure 7, has several specific advantages when used as an implementation platform for business process reengineering:

- since all of the applications pages (e.g., Java applets) are delivered through the same HTTP server and Access Manager they can be either generated dynamically or chosen from the file system accordingly to the process model version in the database. This means that appropriate application or its version can be selected automatically by the system with no or little involvement of the user,

- since all of the major browsers provide built-in authentication mechanisms, this feature in connection with some logic in Access Manger can be used for automatic resolution of the most appropriate application version or configuration for current user - depending on his/her role in the system. For example, different version of a report form can be used by a secretary and different by the accountant officer, although both of them simply select "create report" option from the appli-
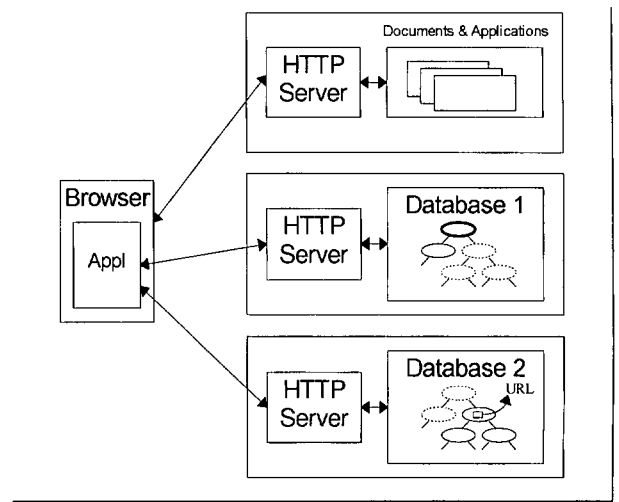
cation menu,

- Access Manager together with a transaction support in the browser (e.g., hidden HTML elements, "cookies", or proprietary Java mechanisms) can be used to determine and impose the appropriate activity sequence in a given business process,

- since the browser is the actual platform for application execution rather than the bare system, it is possible for a user to start more browser instances and work on two or more business processes simultaneously; This feature is very important as the system itself can only schedule actions inside a process, but cannot choose the most appropriate sequence of activities in a set of concurrently executed business processes.

Increasing the model complexity of the business processes in an enterprise (or some other efficiency reasons) may require that the database is distributed among multiple servers. In a big system with great number of concurrently active users such solution can significantly increase the overall performance. With the database model presented in Section 3 and the use of intranet technologies such configuration is legal. An example of such distributed system is presented in Figure 8.

Since, in our model, the communication among database versions is not required, split of the database version three into two or more independent databases is possible when logical addresses of the objects in the database are replaced by the corresponding URL addresses. From the application point of view, in such a case, the whole system is perceived as one distributed database addressed by the use of URLs.

In the example given in Figure 8, the system consists of three HTTP servers. One of them is used only for documents delivery, while two other are connected

to two independent database systems which store two distinct parts of the version tree of the business process models in the enterprise. The first server can be used for delivery of documents and application pages. These pages can be addressed by the application running inside the browser and selected on the basis of user request, authorization data and data contained in currently used database version. During execution applications can also communicate with both of the database systems (addressing them by URL-s) to obtain appropriate versions of the required objects. Addressing of documents, applications, and other objects in the database is accomplished by the use of URL to enable distribution of the data among different servers.

To illustrate how the system works, let us consider two examples. The first example shows how a system administrator can modify an existing process model to reflect a recent change in the operation rules in the enterprise. The second one shows how two employees of sales department use the system do develop a new offer for another company. Our enterprise employs a hypothetical intranet system.

5   Example 1

According to the new instructions given by the enterprise management, all business offers with the value exceeding twenty thousand prepared for the corporate customers have to be approved by the sales manager. Administrator of the information system modifies the system by introduction of a new version of the model of business offer preparation process. The administrator starts new instance of intranet browser and selects "administration" function. A list of few possible options appears one of them being derivation of a new version of an existing process model. After selecting this option the list of all existing process models appears.

Each business process is described in different version of the database. They form a list-like structure since all of them are placed on the same (second) level of database version derivation tree.

After selecting the business process to be modified the system displays the version derivation tree of this process.

Each process model can have multiple versions which form a sub-tree starting from the second level in the version derivation tree.

The administrator selects the appropriate version of the process to be used as a source for the newly derived process version. A new process version appears. On the beginning it is identical to the original version is was derived from.

Creation of a new version of a database involves creation of only logical copies of the objects. Each new version on the beginning contains exactly the same versions of all objects as the parent version.

The administrator modifies the activity graph in this process version by introducing a new activity (review of the offer) which is to be performed by the sales manager. This activity is being inserted between the activities representing preparation of the document by a secretary and sending the document to the director. The result of this activity can be positive or negative. In the first case the document (containing the text of the offer) is being sent to the director. In the second case it returns to the secretary. Finally the administrator modifies the name of both of the process model versions.

All objects which are to be modified in the new database version are first converted from logical to physical copies, therefore modification of these objects does not affect other versions of the database. New database versions are automatically visible for all users which are permitted to use them.

6   Example 2

On the next day a secretary begins work in her office. Already running intranet browser displays a welcome page of the system. She begins her work by entering the main application page. A window with two fields allowing her to enter name and password appears.

Before starting work with the system a user has to provide it his/her authentication data. Since the authentication functions are a built-in feature of the browser, it is automatically performed when the user enters one of the application pages for the first time in the current session. Browser keeps this information to avoid further authentication every time an application page is accessed. After successful authentication, the server reads the user's profile from the database. Since all of the application pages can be generated automatically, their version or configuration may depend on the user's role in the system. Depending on the user's role, appropriate functions are made accessible and current task list is being displayed.

The main system page appears in her browser. This page contains three sections. First section, the biggest, is used for reading and writing documents. Two other sections contain listing of current tasks to be accomplished and the possible actions. There is also a message from the system administrator explaining the most recent changes in the system. After reading this message secretary reads new tasks to be accomplished. In the task window one action is listed: "Write an offer for A Company, Ltd." signed by the sales manager. This task has been already associated by the manager with a generic business process model - "creation of business offers". Since the offer exceeds twenty thousand in value the secretary selects the new of version of the offer preparation process and begins its execution. This process begins with the action of writing a document by the secretary.

Application and its configuration may depend also on the actions previously undertaken and previous activities in the modeled business process. Depending on the option selected by the user, the server sends appropriate document or application page to the client.

In the main section of the client program an editor becomes active and the appropriate document template is automatically loaded. The secretary writes the requested document by filling-in the template. In the meantime, by the use of different browser window, she reads some documents prepared in the technical department to learn about new features offered by equipment being produced. After finishing her work, she saves the document.

On the basis of the previously obtained data, server can automatically prepare appropriate configuration of client application before sending it to the user, e.g., cause the editor to load selected document template. Multiple browser windows may be used to accomplish tasks which can be difficult to foresee on the application development stage and difficult to implement in one application, e.g., reading some documents while writing another one. All documents created by users are maintained by the server. The server controls how the documents circulate in the enterprise. If a document has been prepared as part of a process modeled in the system, it is the server task to send the document to the person responsible for the next stage.

The sales manager receives a message saying that the secretary has finished edition of the document. The sales manager selects this task and the currently default function "read an existing document" function. The document appears in the main window of the browser. He starts reading the document.

The process model stored in the database specifies what actions can be accomplished by users in particular stages of the process.

The sales manager has two options to choose. He may, or may not, accept the current form of the document. Since the offer is lacking some important warranty information, he decides to not accept it.

The next step in the modeled process often depends on the results of the previous stage.

On the secretary's screen new task appears. The offer she had prepared needs some corrections. She receives also a message from the manager explaining what is the concern. She corrects the document. Finally the document returns to the manager, who accepts it this time. The document together with the task description is sent further, to the director.

7 Conclusion

The main goal of this paper was to propose a flexible, persistent and distributed environment for the business process modeling and execution. A special emphasis has been put on the evolution of process models over time which responds to natural requirements of the enterprises, dynamically reacting to the changes at the market, in order to increase or at least preserve the benefits achieved up to date. In the approach proposed in this paper the business process models are no longer static since they are multiversion. Version derivation can be performed in different directions, thus a single business process can have many variant models which reflect alternative ways of process execution.

The proposed approach is based on two modern and very promising information technologies, namely the technology of multiversion object-oriented databases and the intranet technology. The database technology, on one hand, provides advanced mechanisms for both storing and accessing persistent objects, like concurrency control, integrity control, access authorization, data recovery. These mechanisms are very useful for supporting enterprises in their daily work, since almost every business process requires efficient storage and retrieval of consistent information. On the other hand, the database technology provides object versioning and the configuration management. These mechanisms make it possible to represent both the different variants of the same process and its revisions in the information system.

Intranet, being the second base technology of the proposed approach, enables practically unlimited flow of information between all employees of an enterprise, and access to all information resources of the enterprise, including historical information. The same concerns hardware and software systems which can be integrated with intranet in a way guaranteeing their full accessibility. Again, this supports the evolution of enterprises which can extend their information and computing platforms in a progressive way, adding new resources and technologies, instead of performing revolutional changes aiming to replace one platform by another.

The resulting integrated platform can present a new value for the enterprise. On one hand, it naturally implements theoretical approach to business process modeling and reengineering, which would be very difficult without intranet technology. On the other hand, it extends the intranet concept with a new functionality, directly addressed to support the enterprise way of work. Finally, the integrated platform offers decision support to the enterprise management, coordination of collaborative work of the enterprise employees, and - what is even more important - an unlimited access to the legacy systems. As a consequence, it can significantly increase efficiency of the enterprise and its competitiveness on the market.

Future work will mostly concern extensions of the proposed approach towards strategic analysis and quality control of processes. Since the evolution of business process model is not random, but it is strictly related to the assessment of the quality of process

model versions already elaborated and executed, there is a natural requirement to assist this evolution by system mechanisms which can support the derivation of new process model versions, relating to the complementary analysis of revisions.

References

[1] Ahmed R., Navathe S.B. (1991) Version Management of Composite Objects in CAD Databases, *Proc. SIGMOD*, Denver, CO, p. 218-227.

[2] Aiello L., Nordi D., Panti M. (1984), Modeling the Office Structure: A First Step Towards the Office Expert System, *Proc. of 2nd ACM SIGOA Conf.*.

[3] Brierley E. (1993) Workflow Today and Tomorrow, *Proc. of Conf. on Document Management.*

[4] Bullinger J.H., Mayer R. (1993) Document Management in Office and Production, Nachrichten fur Dokumentation, Vol. 44.

[5] Cellary W., Jomier G. (1990) Consistency of Versions in Object- Oriented Databases, *Proc. 16th VLDB Conf.*, Brisbane, Australia, 1990

[6] Estublier J. and Casallas R. (1994) The Adele Configuration Manager, *Configuration Management*, ed. by W. Tichy, Wiley and Sons, Software Trend Series.

[7] A.J. Freed (1997) Net Asset: Intranets Have Much To Offer Businesses, http://www.dciexpo.com-/news/archives/intranet.htm

[8] Hendley T. (1992) Workflow Management Software, Information Management and Technology, Vol. 25.

[9] Jones J.I., Morrison K.R. (1993) Work Flow and Electronic Document Management, Computers and Industrial Engineering, Vol. 25.

[10] Kay M.H., Rivett P.J., Walters T.J. (1992) The Raleigh Activity Model : Integrating versions, concurrency, and access control, *Proc. BNCOD.*

[11] Kofer W., Schoning H. (1992) Mapping a version model to a complex- object data model, *Proc. Data Engineering*, Tempe, AZ.

[12] Sarda N. (1990) Extensions to SQL for historical databases, IEEE Transactions on Knowledge and DataEngineering, vol. 2, Nr. 2, p. 220-230.

[13] Tansel A. U., Clifford J., Gadia S., Jajodia S., Segev A. and Snodgrass R. (1993) *Temporal Databases: Theory, Design, and Implementation*, Benjamin / Cummings, Database Systems and Applications Series.

[14] Talens G., Oussalah C., Colinas M.F. (1993) Versions of simple and composite objects, *Proc. 19th VLDB*, Dublin.

[15] Wieczerzycki W. (1996) Process Modeling and Execution in Workflow Management Systems by Event-Driven Versioning, *Business Process Modelling*, eds. B. Scholz-Reiter, E. Stickel, Springer-Verlag, p.43-66.

[16] Wieczerzycki W., Rykowski J. (1994) Version Support for CAD/CASE Databases, *Proc. of the East-West Database Workshop*, Springer-Verlag.

[17] *The Intranet Implementation of Internet And Web Technologies In Organizational Information Systems* (1996) Hummingbird Communications Ltd., http://www.hummingbird.com-/whites/intranet.html

[18] *The Intranet Definition, Benefits and Challenges* (1997) http://pscinfo.pscni.nasa.gov-/intranet/html/home.html

[19] *Netscape Intranet White Paper, Intranets Redefine Corporate Information Systems* (1997) http://partner.netscape.com/comprod/at_work/-white_paper/indepth.html

Software for Constructing and Managing Mission-Critical Applications on the Internet

Piotr Dzikowski
Talex Sp. z o.o., ul. Pułtuska 10, Poznań, Poland
Phone: +48 61 8792901 ext. 112
E-mail: `piotrd@talex.com.pl`

This paper reviews the technology, customer requirements, and architecture behind BEA Jolt and BEA TUXEDO - the products which can enable users to bring new or existing applications to the Internet almost immediately, using the expertise they already have, with the reliability and scalability that mission- critical applications demand. This allows users to use both complex client/server and commercial on-line transaction processing (OLTP) applications with scalability, flexibility, and maintainability to stand up to the rigorous requirements of today's information-driven business world.

1 Introduction

The growth of the Internet is providing business with new ways to interact with current and potential customers. Many companies have already jumped at the opportunity and are testing sample applications on the Internet (via the World Wide Web) that let customers gather basic information and, in some cases, initiate simple buy/sell transactions. So far, however, the true business value of the Internet as a ubiquitous communications medium remains mostly unrealized. While current technology allows some basic business to take place over the Internet, organizations are eager to explore the more complex and much more lucrative possibilities of mission-critical Internet commerce. In fact, whether mission-critical applications are isolated within a company's internal information systems or allow access from the outside via the Internet, they all have very similar requirements. Applications in both instances require transaction and data integrity, reliability, scalability, and security. The Internet, however, is outside of company's realm of control, making management and administration significant issues. The he protocols that the Internet utilizes, have the additional handicap of being unable to recognize the state of a transaction, making the complex multi-step interactions, necessary for the mission-critical Internet commerce, a technical challenge (see Figure 1). Mission-critical Internet commerce requires a whole new level of application infrastructure to overcome these hurdles. It requires the infrastructure to support the integrity of the back-end servers, plus a new appliance of this infrastructure built specifically to bring it safely and reliably to the Internet, while ensuring integration with legacy applications.

BEA Systems Inc., provides its BEA Distributed Application Framework - with the BEA TUXEDO transaction and messaging engine at its core - as the robust middleware infrastructure needed to develop and deploy mission-critical applications. In response to the significant customer interest in the Internet commerce, BEA is now also providing the BEA Jolt software package to connect this middleware framework to the Internet, enabling BEA's customers to place themselves in the middle of the new business opportunities.

2 Brief overview of the TUXEDO system

The TUXEDO system provides an industry standard for creation and central administration of distributed on-line transaction applications in a heterogeneous client/server environment. It allows to create and to maintain reliable, high performance, easily managed distributed systems. The Application Administration functions are primarily concerned with management of user-created components such as:

- Domains

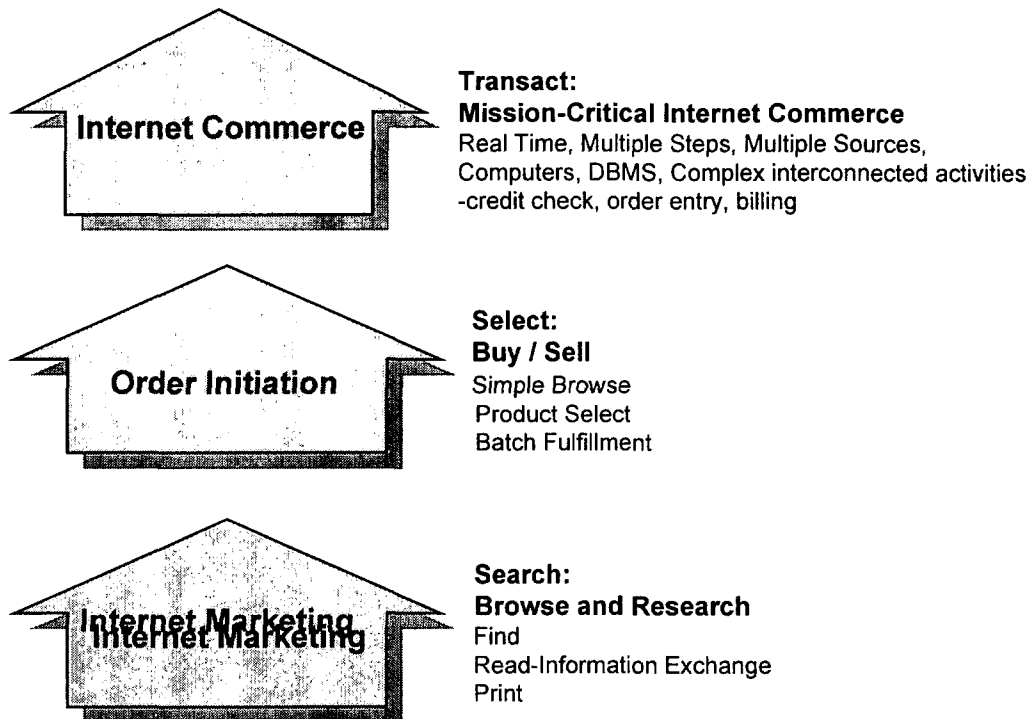- Servers (Hardware and Software)

- Clients

- Queues

Transact:
Mission-Critical Internet Commerce
Real Time, Multiple Steps, Multiple Sources,
Computers, DBMS, Complex interconnected activities
-credit check, order entry, billing

Select:
Buy / Sell
Simple Browse
Product Select
Batch Fulfillment

Search:
Browse and Research
Find
Read-Information Exchange
Print

Figure 1: The three levels of Internet commerce each require a different technical infrastructure. Mission-critical Internet commerce consists of multiple, complex transactions and requires a robust middleware infrastructure with the same reliability, scalability, and security as internal mission- critical applications.

– Groups

– Services

Thanks to the TP monitor it is possible to manage [2]:

– application messaging to provide transactional reliability and integrity,

– sharing components by other application entities,

– recovery of failed messages,

– scheduling of communications.

3   The Managed Multi-Tier client/server model

The TP Monitor manages the data flow, so servers can receive requests from multiple clients. There does not have to be a one-to-one relationship between clients and servers. This also means that clients can make requests on multiple servers. This feature combined with TP Monitor's transaction management, provides a powerful Distributed Transaction Processing (DTP) framework.

In the Managed Multi-tier (MMT) C/S model, a Transaction Processing (TP) monitor is introduced to provide:

– communication between the client and the server,

– a method for powerful system administration,

– a method for transaction management.

Since the TP Monitor assumes control of both communications and transaction management, the database engine can concentrate on what it does best: managing the data! In this model, the database becomes a pure resource Manager (RM) which is a generic abstraction for any managed transactional entity coordinated by the TP Monitor.

4   Overview of BEA Jolt

4.1   Introduction

BEA Jolt takes mission-critical enterprise applications to the Internet by extending access for both new and current corporate applications to wherever Java can be run. BEA Jolt is a set of software components to let Java programmers make BEA TUXEDO service requests from the Java language [5,6]. BEA Jolt supports both Internet and Intranet development and deployment, operates safety through firewalls, can be used to access mainframe applications, and provides a sound methodology for building robust, scaleable, mission-critical Internet applications (see Figure 2).

Tools for developing Internet applications are now emerging, focusing on building screens and accessing database files. Such tools do not, and cannot hope to, build scaleable, distributed, mission-critical applications without a supporting infrastructure. The BEA Distributed Application Framework provides such support through a suite of integrated products, of which BEA Jolt is the newest components.

4.2 Background

Understanding the benefits of BEA Jolt requires an understanding of how the Internet systems are constructed today, the role of the Java-enabled Web browses, and how these technologies relate to mission-critical systems. Mission-critical systems are those where reliable, predicable performance meeting specific criteria can be expected with certainty. The use of the Internet as a sales and customer care channel is beginning the transformation of the requirements of Internet applications from novelties and experiments to mission-critical systems. BEA Jolt directly addresses the issues discussed above [4].

4.3 The BEA Jolt Architecture

BEA Jolt connects the Java clients to the BEA TUXEDO applications. A BEA TUXEDO application is a set of services, each offering some specific functionality related to the application as a whole. A simple banking application might have services such as Inquiry, Withdrawal, Transfer, and Deposit. Normally, such service requests are implemented in C or COBOL as a sequence of calls to a program library. In the case of Java applets, access to such libraries is prohibited. Even if it was allowed, accessing a library from a Java program means installing the library for the specific combination of CPU and operating system release on the client machine, a situation that Java was expressly designed to avoid. Instead, Jolt provides an alternate implementation of the TUXEDO client library partiularly suited for the requirements of the Java applets. This implementation acts as a proxy for the native TUXEDO client, implementing the functionality available via the native TUXEDO client.

The Jolt server accepts the requests from the Jolt clients and maps those requests into the TUXEDO service requests via the TUXEDO ATMI API. The request and the associated parameters are packaged into a message buffer and delivered over the network to the BEA Jolt server. The Jolt server unpacks the data from the message, performs any data conversions necessary, such as numeric format conversions, character set conversions, etc., and makes the appropriate service request to the TUXEDO as specified by the message. Once a service request enters TUXEDO, it is executed in exactly the same manner as any other
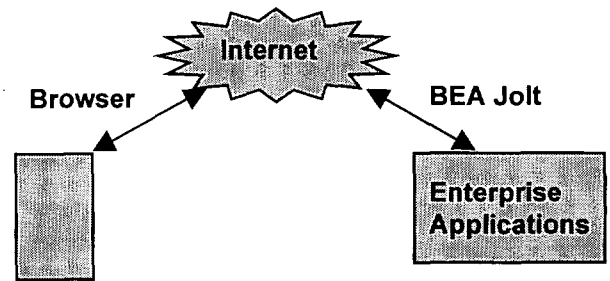

Figure 4: Internet Application Topology.

TUXEDO request, and the results are returned via the ATMI interface to the BEA Jolt server, which packages the results and any error information into a message which is sent to the BEA Jolt client code. The Jolt client then maps the contents of the message into the various objects comprising the Jolt client interface, completing the request. Jolt is fully compatible with Java threads, simplifying program design and implementation (see Figure 3).

Most importantly, Jolt works with the standard TUXEDO. Current application servers serving the enterprise customers can be Jolt-enabled without a single change to the application, allowing application developers to focus on the high level policy issues associated with using an enterprise application with Internet, not on the programming details of the interface. An important advantage of TUXEDO is that its administration capabilities are very mature and robust. Unlike monolithic applications, which require programming changes to implement the access policies, TUXEDO provides the on-line administration tools that control access to the individual TUXEDO services. This means that bringing the enterprise to the Internet is much more a matter of an administrative configuration than programming, shortening the development cycle and increasing the confidence in the safety of the applications presented to the Internet.

Of course, Jolt addresses both the Internet and Intranet application development. Figure 4 shows an overview of various hardware components that make up an application: the Internet access, the network security devices and the administrative control. Internet applications require access security, the method of choice being the installation of firewall systems. Intranet applications typically never encounter a firewall, as corporate network policies usually presume a level of trust of corporate users orders of magnitude higher than those of the Internet users [6].

Jolt provides means for safe access to applications through the corporate firewalls. The currently standard release of Java for programmers, the Java Development Kit release 1.0 (JDK 1.0), does not provide
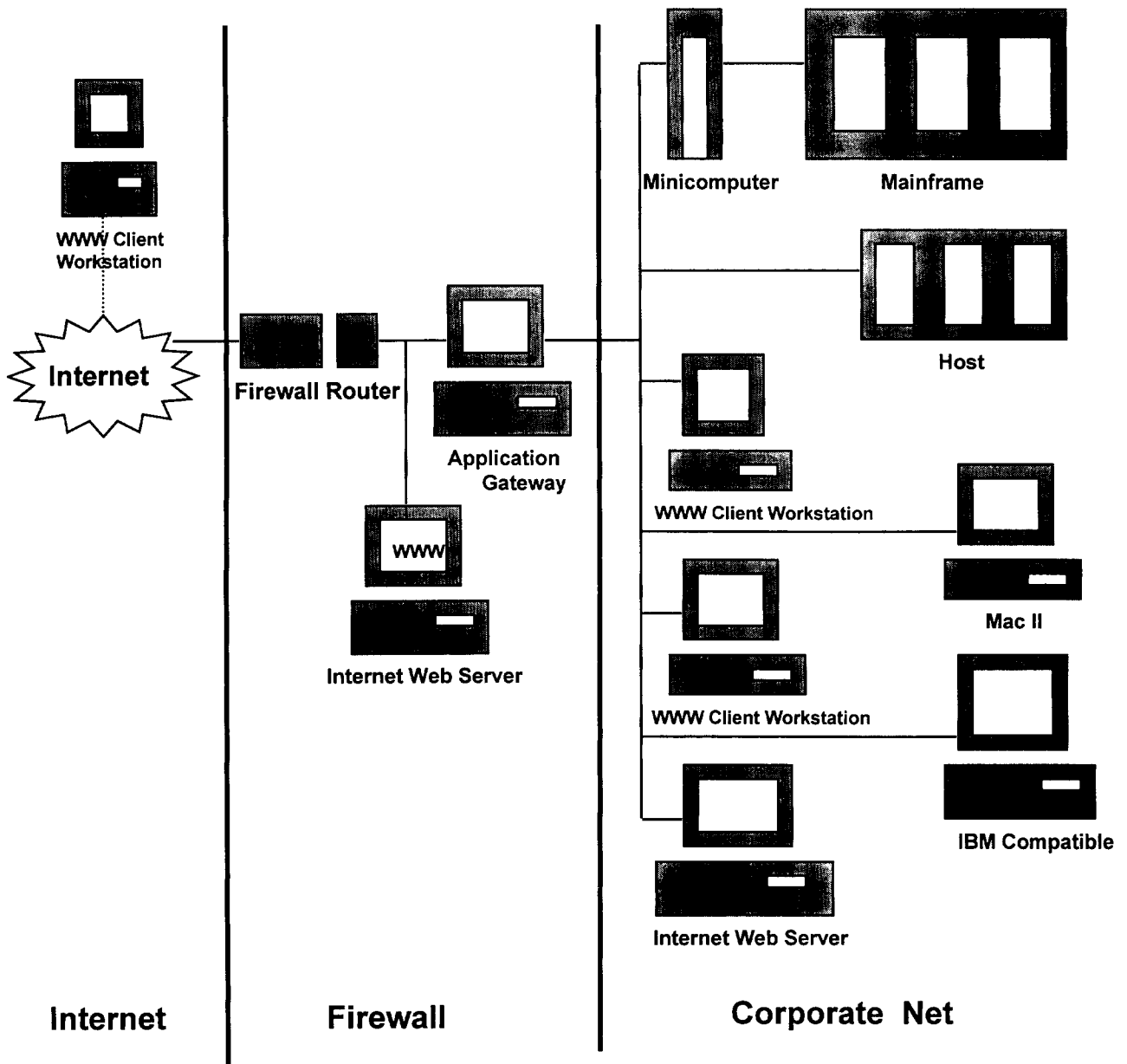
Figure 2: BEA Jolt enables mission-critical enterprise applications to have turn- key access to the Internet.
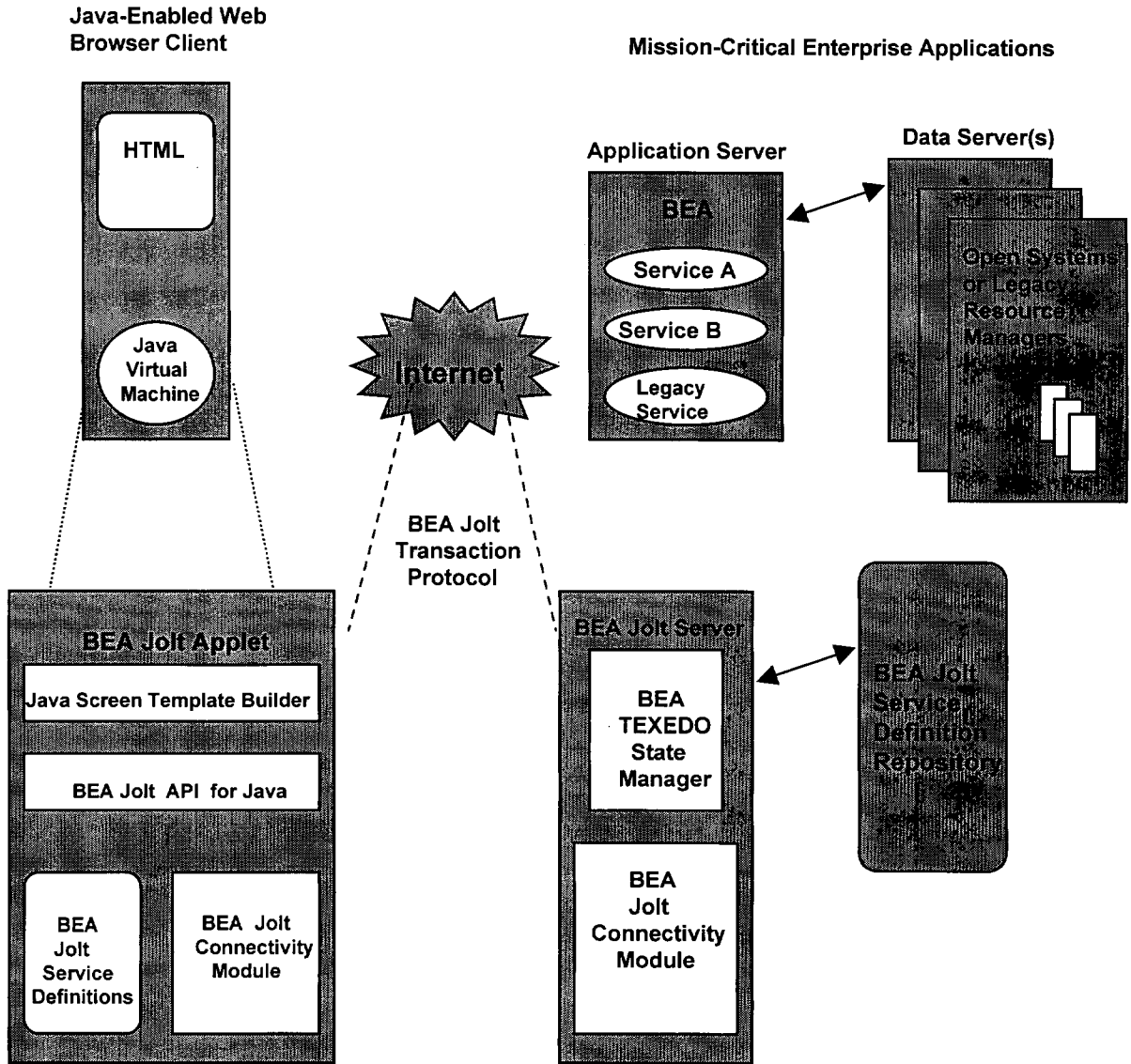
Figure 3: The BEA Jolt architecture is created to enable Java programmers to build graphical front-ends that utilize the application and transaction services of the BEA TUXEDO without needing to send detailed transactional semantics over the Internet.

many important facilities necessary to implement secure access, including access to SSL for transmission security and signed applets. (Signed applets use public key encryption technology to encrypt a Java applet in such a way that authenticity can be assured to an extremely high level of confidence.) BEA Jolt will provide scalable access through firewalls and work with forthcoming Java security features.

4.4   BEA Jolt and the Web

While popular, CGI programming is very resource intensive compared to the demands of transaction processing, where reducing every extra cycle may result in an increased throughput. Jolt takes advantage of Java by avoiding use of the CGI programs entirely. It uses the WWW to download the Java applet using this product embedded in an HTML page. Once the page is loaded, the Web browser activates the applet, which then becomes an autonomous program running in the browser. Rather than attempt to communicate with the TUXEDO services through the Web server, Jolt opens a direct connection to the BEA Jolt server. There are several advantages to this approach. First, communications with the Web servers require use of protocols understood by the Web server, most notably http. Encoding a service request as an http request involves considerable extra coding in Java and much more overhead than the simple lightweight protocol that Jolt employs. Second, an http request must flow through the Web server, and either activate a CGI program to handle the TUXEDO service request or speak to a companion program attached to the Web server via NSAPI or ISAPI. While these direct attachments greatly improve performance, the Web server still must process the request. The companion program is essentially the interface to the TUXEDO, which is exactly the function of the BEA Jolt server. Rather than introduce the Web server as an unnecessary intermediary, the direct connection is employed. Third, the Web server may in fact limit the scalability of the application. Any limits imposed by the Web server, such as the maximum number of concurrent connections, maximum number of sockets per process, etc, will limit the access of Jolt applets to the application framework (see Figure 5). Web servers attempt to make the best use of the limited network resources by opening a connection from the browser to the server for the duration of the download of a single resource. A Web page which includes references to 10 GIF images will require 11 separate connections, one for the page, and another for each GIF. This is akin to opening and closing a file every time a single line is read. Jolt provides an optimal network resource usage by making the duration of a connection a policy choice by the application. A connection can be open for as little as a single request or as long as the activation

time of the applet. For applets which are performing many service requests, long term connections are much more efficient than that of Web servers. In addition, application state is maintained even when the network connection is released. An individual applet can be identified uniquely over many different service requests, regardless of whether the network connection is maintained or re-established each time. Such identification greatly simplifies the structure of the Jolt application development over that of CGI-based programs, where such identification is problematic at best. Finally, there is no indication that the Web is the final product of the Internet product evolution. Implementing Jolt independently of the Web means it can more easily accommodate the new Internet technologies as they become popular. All this being said, it is also possible to use Jolt through a Web server as described above, for those who have the special requirement to do so.

4.5   Scalability

Success of the Internet applications is usually measured by the number of users attracted to an application over a period of time. The goal is to service as many users as possible with the fewest number of problems. The term usually applied to this goal is scalability, and means that incremental requirements can be met by incremental resource additions. Adding support for 10 more users should not require a system redesign, a network overhaul, or re-implementation of the database.

Building scalable systems requires a sound architectural foundation and consistency of design and implementation. Each component of a system must by capable of meeting the demands placed upon it in with no revolutionary architectural or technological changes. For Internet systems, there are three vital areas of scalability: database, application, and network.

By using TUXEDO, to balance loads and coordinate transactions between multiple servers, database scalability has been dramatically improved [2,3]. Application scalability is easy to overlook and difficult to correct without a rewrite. Often the desire to provide applications quickly interferes with the ability of the application to effectively access corporate data without interfering with other users, due to database locking, CPU, or I/O resources. Such problems are typical for monolithic applications, where all of the various subcomponents of an application are combined into a single, large program. Using a middleware infrastructure engine like TUXEDO allows an application to be split into a set of services that communicate with messages enabling individual tuning, performance improvement, and access control of any of the component. Application scalability provides the control necessary to reconfigure the resources through the
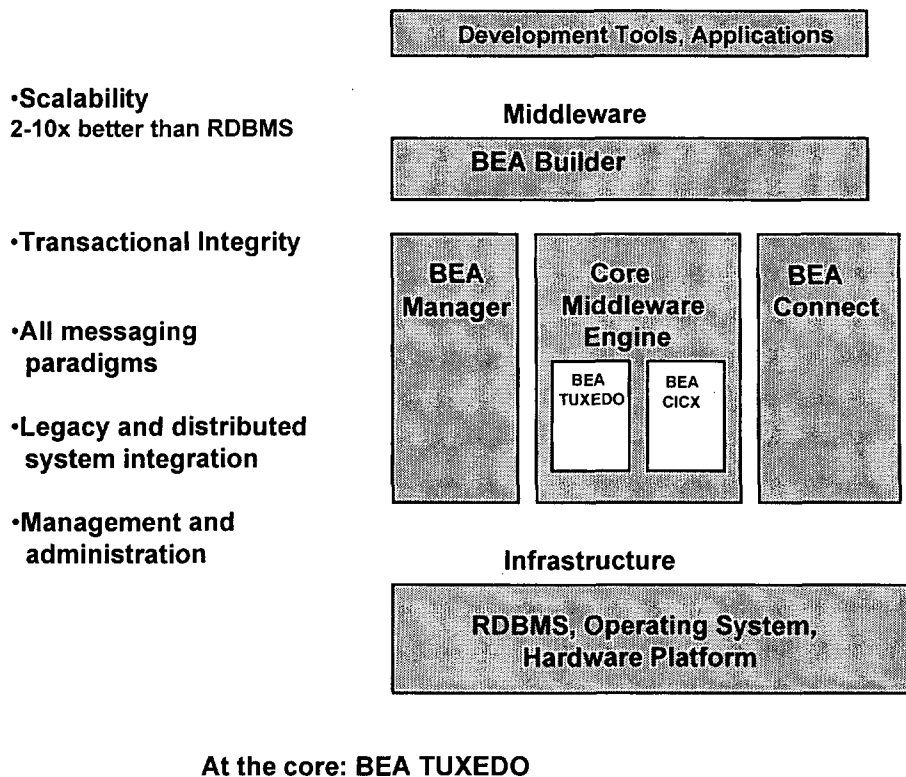
·Scalability
2-10x better than RDBMS

·Transactional Integrity

·All messaging
paradigms

·Legacy and distributed
system integration

·Management and
administration

Development Tools, Applications

Middleware

BEA Builder

BEA Manager	Core Middleware Engine	BEA Connect
	BEA TUXEDO	BEA CICX	

Infrastructure

RDBMS, Operating System,
Hardware Platform

At the core: BEA TUXEDO

Figure 5: BEA Jolt brings the benefits of the BEA Distributed Application Framework to the Internet.

administration, not re-implementation.

Finally, being able to modularize the use of the network resources is an important aspect of the Internet application operation. The introduction of the new networking facilities, or modifications to the existing installations, should only involve some administrative changes to the application. The firewall negotiation facilities of BEA Jolt provide scaleable concentration of multiple network requests into a smaller set of request streams, easily and safely routed through application gateways and corporate systems. Multiple firewall negotiations can be used to accommodate very large network installations. The load balancing capabilities of TUXEDO are employed to ensure that high throughput attends high request volume. Jolt works with TUXEDO to simplify scalability by providing a modular framework providing a network, application, and database access scalability [1].

5   Conclusions

This paper was written to convince that MMT C/S Architecture and Internet made by BEA Tuxedo and BEA Jolt is the way to go. The reason for choosing a client/server and Internet over a mainframe solution is in the mainframe's inflexibility. Business reengineering has been the Achilles heel for mainframe technology,

not because it's expensive and requires a lot of people to operate. It's worse; it does not lend itself to even minor changes, let alone the reengineering of an entire enterprise. So businesses driven by mainframe technology find it nearly impossible to respond to ever changing market conditions in a timely fashion.

The main goal of this paper was to introduce a totally new solution, for an information system that would take companies 10 to 15 years into the future, and help them to surround their business applications with the latest technologies. The author of this document is a employee of TALEX company which provides its customers with both medium and large-scale distributed, on-line transaction, business applications spanning as different platforms as UNIX Sun Solaris, SCO Open Server, HP, NT, Windows 95, DOS with both GUI and text-based support. He is responsible for updating new technologies at TALEX.

References

[1] BEA Systems, Inc. (1996) *Manuals.*

[2] C. L. Hall (1996) *Technical Foundations of Client/Server Systems.*

[3] X/Open Company Ltd. (1993) *Distributed Transaction Processing: The XA Specification*, Berkshire, UK.

[4] H. Beacon, (Winter 1996/97) *BEA 'JOLTs' the Internet*, Keeping Bea customers informed.

[5] BEA Systems, Inc. (1996) *Bea Jolt - BEA Enterprise Middleware Solutions.*

[6] BEA Systems, Inc. (1996) *Bea Jolt White Paper.*

The New Software Technologies in Information Systems

Dariusz Smoczyński
TALEX-PSO, ul. Lacina 1, 61-132 Poznań, Poland
E-mail: dareks@talex.com.pl

Information systems used in many organizations are wholly based on a single relational database technology such as Informix or Progress. The emerging of new software technologies makes it possible to add more functionality to working database applications. The question is what solutions to choose? As a software company we see that the most promising technologies seem to be intranet, workflow, SGML and middleware.

1 Introduction

Talex is a software company aimed at developing relatively small information systems with up to 40 users. Our customers are ready to accept some costs of maintaining such systems at the level of 2–3 programmers. This means that we while we can encounter problems typical for large systems we cannot afford using advanced software products (like IBM CICS). In our information systems we would like to offer solutions that represent the best price/performance ratio while they are developed using the simplest software tools. Our efforts are inspired by the Donald E. Knuth's typesetting system TEX. Instead of fancy features the TEX system is:

- reliable

- inexpensive

- open for easy modifications and improvements

The same goals we would like to achieve building our information systems.

2 Multi Server Information System

The model of a multi server information system is shown in Figure 1. The term *multi server* is to stress that the system is based on different software tools which are combined to create the complete system. The multi server information system is accomplished in phases from the existing single database application. There are three main goals when new software technologies are introduced to the working database application:

1. to process data stored in the existing database in many new ways to satisfy the personal needs (for example to apply Microsoft Excel to the existing data),

2. to store and process new types of documents (e.g. text or images) which are difficult to maintain by relational database,

3. to integrate database applications with other systems like data warehouse.

As shown in Figure 1 there are three layers in the proposed multi server information system model: personal desktop layer, server layer and information resources layer.

Personal desktop is a set of software tools to support any data processing needed by a specific user. As a good example we can mention Microsoft Office package. Using MS Excel, MS Word or MS Power Point one can obtain an almost complete functionality of data processing. One should notice the key role of the Web browser (in Figure 1 it is the Microsoft Internet Explorer). It is a gateway between the other desktop applications and the information resources.

Server layer creates a program environment for the applications executed on the personal desktop. The environment supplies the following services:

- Data manipulation services (select, insert, delete, update) for maintaining the data from information resources layer.

- Data dictionary for the help systems.

- Log facilities to trace all system activities. The log is used for the system optimization and error recovery.
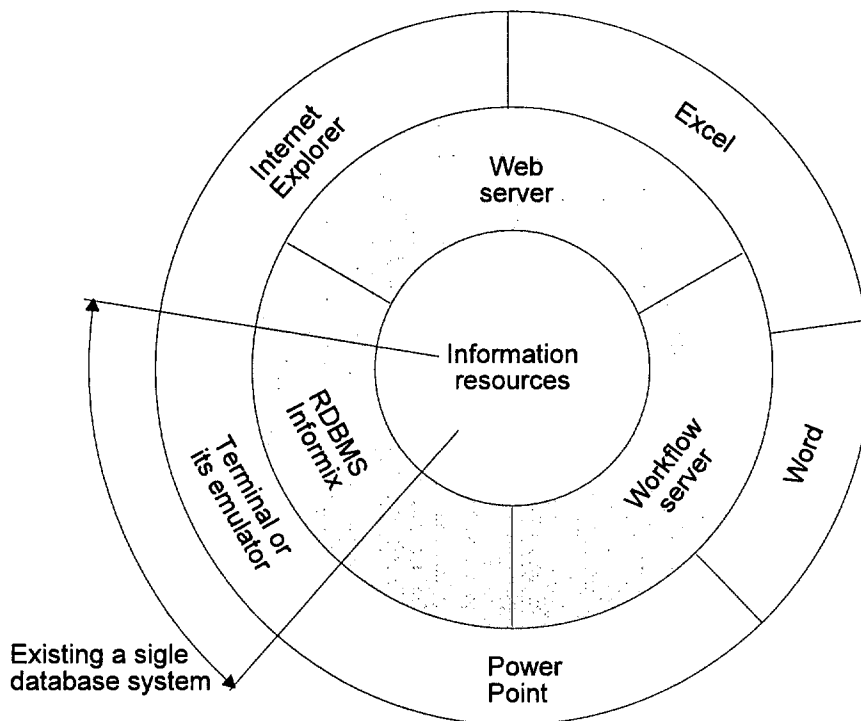
Figure 1. Multi Server Information System . A set of various servers enables information sources to be used by applications from personal desktop layer.

- Access Authorization for applications and data.
- Workflow services.
- SGML backup services for platform independent storing of data.

Information resources consist of all documents used in an organization.

We will now describe the place and role of the previously mentioned software technologies: intranet, workflow, SGML and middleware in the proposed multi server information system.

3   Intranet

Intranet allows getting access to the information resources using the Web servers and browsers. As a part of the server layer in the multi server information system model, Figure 1, WWW server supplies the data manipulation services.

WWW server is able to deliver to the client browser any data type executing Common Gateway Interface

(CGI) programs. The structure of the Intranet is shown in Figure 2:

The CGI program establishes a connection to a database server, selects data and returns it as a dynamic HTML page. The page is then sent to the browser on the personal desktop layer. In our information system we have three types of CGI:

CGI type H converts SQL data to the dynamic HTML page. Such pages are intended to serve as reports because they can be printed using the Web browser facilities. You can also cut and paste any piece of information from the browser window to any other application for instance MS Word.

CGI type D converts SQL data to DBF format and creates a dynamic HTML page with the FTP service. Web server transfers the DBF file to the workstation and then the file can be imported for instance to the MS Excel.

CGI type V calls or first creates and then calls Java applets.
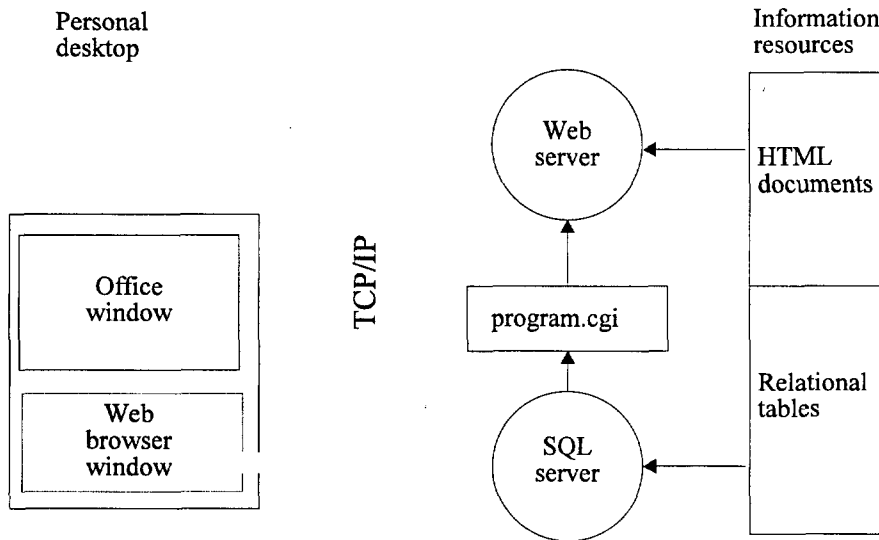
Figure 2. CGI program creates dynamic HTML page and enables any type of information to be delivered by the Web server to any application on personal desktop.

4   Workflow

Workflow management is the automated coordination and control of work processes. As a part of the server layer in the multi server information system model, Figure 1, workflow server coordinates and controls the pieces of software on the personal desktop layer.

Workflow management systems are offered by many vendors. IBM FlowMark can be an example. The Workflow Management Coalition (WMC) is an international organization for the development and promotion of the workflow standards. In our project we implemented a server called the *dynamic application server*. The server is based on WMC specifications but is only a subset of a full scale workflow server. The dynamic application server controls the execution of **only our own** application programs. The architecture of the server is shown in Figure 3:

5   SGML

Standard Generalized Markup Language (SGML) is an ISO standard for defining the document structure. It enables the platform independent data description.

Periodical backups created by a database utility are not very easy to use by another system. Even if you are using the same database for a long time you may have a problem to restore some data if in the meantime your database schema has been changed. In this context we must remember that some data (e.g. financial) should

be kept almost forever. There is a problem how to store the data to be independent of any platform and be available to be used at any time.

The SGML-encoded documents are standard ASCII files containing *tags* and *data items*. Tags determine the logical structure of the data. In the following example the tag <from> defines the data item containing the name of a sender:

```
<to> all programmers
<from> Jan Kowalski
<date> 12th December 1996
<subject> SGML seminar
<para> We meet on the next Monday...
```

SGML tags must be defined in another ASCII file called the Document Type Definition (DTD). Having both SGML-encoded document and its DTD allows for computer to make any transformation of the data to meet the requirements of the specific system.

5.1   Forever green archive

SGML technology can be used to make platform independent archive which we called the *forever green archive*. Its model is shown in Figure 4:

The idea of the archive is to create the SGML-encoded documents starting from reports produced by the application programs. The main advantage of our approach is that reports represent the human view of data. The alternative would be to create an archive directly from relational tables described by the SQL statements.
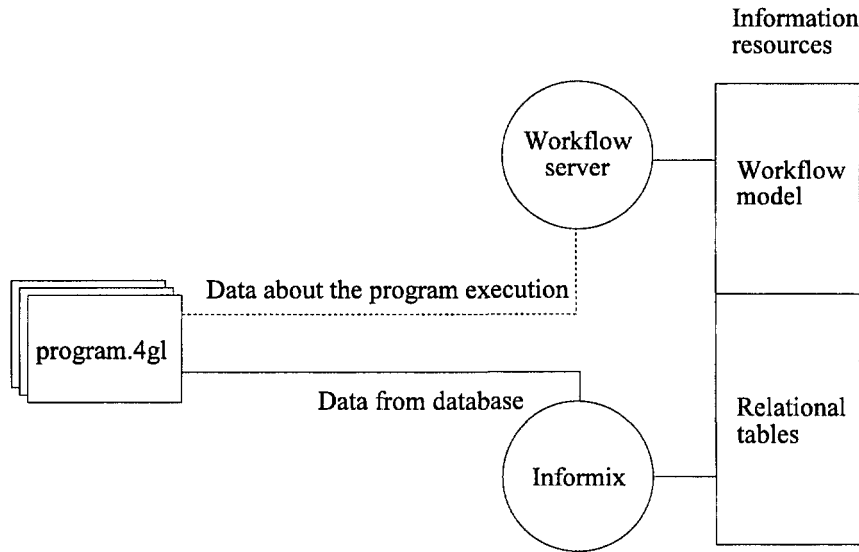
Figure 3. Workflow management system controls the execution of application programs. Workflow server can do its tasks using its own repository.
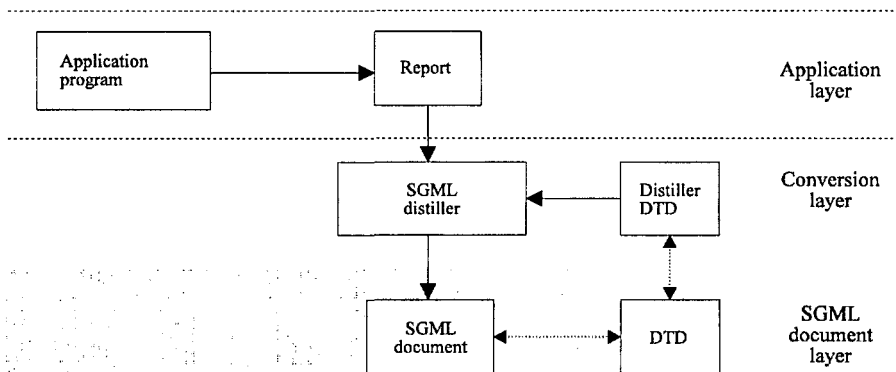

Figure 4. Creating SGML-encoded documents and platform independent archive for information system. The source of data is taken from a report.

6   Middleware

Middleware is a special software layer for hiding the server diversity and enable the personal desktop applications to see only one virtual server. Middleware approach enables applications to be independent of physical structure of the system.

The use of the middleware changes client/server architecture to the three-tier client/server model as shown in Figure 5:

Applications that form the personal desktop layer interact with the user. The interaction is shown at Figure 5 as the INPUT and DISPLAY statements. Database access is performed by other piece of software – middleware – via the CALL statements. This middleware approach enables the database stucture to be changed without affecting the application programs.

6.1   Multi Tier Application Builder

Using the middleware benefits requires an appropriate structure of the application programs. We have developed a multi-tier application builder KOMETA to simplify the task of bulding and maintainig the multi-tier applications. KOMETA creates programs which structure is shown in Figure 6:

In the KOMETA system we can distingush:

driver   a main module created by the KOMETA builder. It is a "black box" – there is no need for an application programmer to change it.

exits   KOMETA builder which creates a set of void procedures e.g. INIT, BEFORE INPUT, AFTER INPUT etc. Application programmer can change the body of these procedures and then tailor the behaviour of an application.
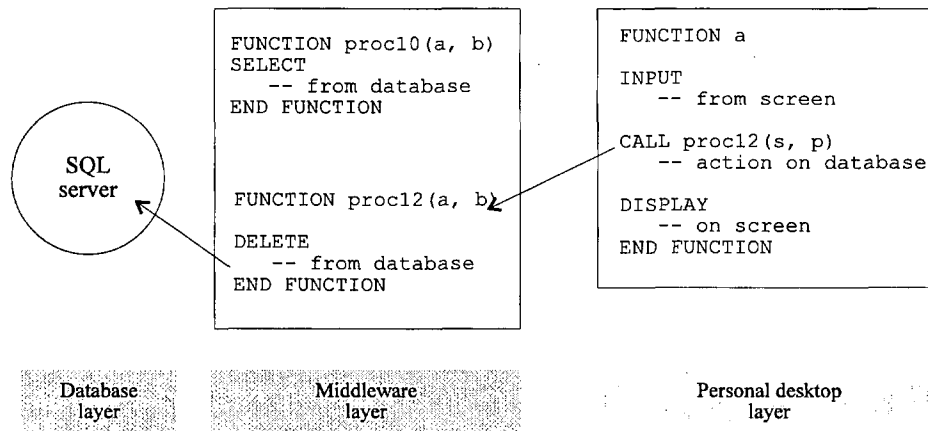
KOMETA API   KOMETA Application Program

```
FUNCTION proc10(a, b)
SELECT
    -- from database
END FUNCTION


FUNCTION proc12(a, b)

DELETE
    -- from database
END FUNCTION
```

```
FUNCTION a

INPUT
    -- from screen

CALL proc12(s, p)
    -- action on database

DISPLAY
    -- on screen
END FUNCTION
```

SQL
server

Database
layer

Middleware
layer

Personal desktop
layer

Figure 5. Multi-tier achitecture enables application programs from personal desktop layer to be independent from physical structure or address of data.
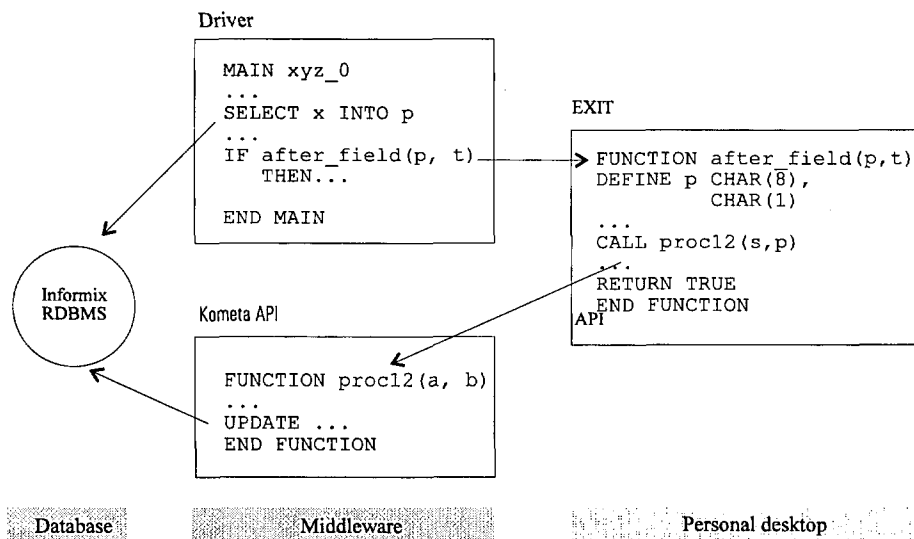
Driver

```
MAIN xyz_0
...
SELECT x INTO p
...
IF after_field(p, t)
    THEN...

END MAIN
```

EXIT

```
FUNCTION after_field(p,t)
DEFINE p CHAR(8),
           CHAR(1)
...
CALL proc12(s,p)
...
RETURN TRUE
END FUNCTION
```

API

Informix
RDBMS

Kometa API

```
FUNCTION proc12(a, b)
...
UPDATE ...
END FUNCTION
```

Database

Middleware

Personal desktop

Figure 6. Kometa application builder unifies all programs structure. It makes more easy to maintain and develop application software.

Interface consists of a set of standard procedures (e.g. financial) which can be used for creating applications.

7   Conclusion

The article presents a model of a multi server information system. It is assumed that the system is to be created step-by-step from a single database system. The Web servers for accessing the different types of information resources, workflow servers to control the execution of applications, SGML standard for achieving data independence and middleware for keeping application software under control are the key components of the system in question.

References

[1] Goldfarb C. F. (1994) *The SGML Handbook*, Clarendon Press, Oxford, UK.

[2] Kador J. (1996) The Ultimate Middleware, Byte, April, p. 79-83.

[3] Rymer J. R. (1996) The Muddle in the Middle, Byte, April, p. 67-70

[4] Salamone S. (1996) Middle(ware) Management-Byte, April, p. 71-77

[5] Workflow Management Coalition (1997) *Interface 2. Application Programming Interface (WAPI) — Specification*, Document Number WFMC-TC-1009

Communication Satellites, Personal Communication Networks and the Internet

Hugo Moortgat
Business Analysis and Computing Systems
San Francisco State University
San Francisco, California 94132
E-mail: moortgat@sfsu.edu

This paper investigates the use of communication satellite systems in implementing two different types of services: Personal Communication Networks and Broadband Internet. It first introduces the main objectives of Personal Communication Networks and explains why wireless systems are best suited to satisfy their requirements. Next, it motivates the idea that the future Internet is to be a broadband network with bandwidth on demand and quality-of-service guarantees. Further, it reviews the principal characteristics of geostationary, low earth orbit, and medium earth orbit satellites and satellite systems. Finally, it presents some key commercial systems which are currently being deployed or are in an advanced design stage and will provide wireless Personal Communication Networks or broadband Internet service on a global scale.

1 Introduction

Telecommunication technology, in all its forms, is developing at a seemingly ever increasing rate. This is to a considerable extent due to the phenomenal development of digital computers in the past decades, which has allowed all forms of communications to switch from analog techniques to the digital ones.

The power of current microprocessors, whether embedded in devices or as parts of more general purpose computer systems, are making it possible for the telecommunication engineers to develop systems whose functionality could not possibly have been considered even a short time ago. Among these systems are current generations of mobile telephone systems and a range of nomadic computing services that are planned to be available in the relative short term. Although satellites have not traditionally been considered to play a prominent role in the framework of Personal Communication Networks, this may very well change with the new mobile satellite communications systems whose first satellites have been launched in the past few months and which are scheduled to become operational in the next few years.

The Internet is experiencing a phenomenal growth. Conventional wisdom says that the proper medium with which to build the Internet infrastructure is optical fiber. But satellite systems which are now on the drawing boards intend to challenge this wisdom. Some of these systems are scheduled to go into operation in the next five years or so.

In this paper we look at communication satellite systems in relation to the Personal Communication Networks and the Internet. We first briefly define some of the key aspects of Personal Communication Networks and motivate the use of wireless systems for them. Next, we make some general observations about the Internet and discuss some drawbacks as well as advantages of a possible satellite-based Internet. We then take a look at communications satellite systems in general. Finally, we review some of the current development activity in satellite systems and illustrate how some make it possible to extend Personal Communications Networks on a global scale and how others aim to offer global Internet functionality and other broadband communication services.

2 Personal Communication Networks

In very broad terms, the objective of Personal Communication Network systems [1,8] is to provide ubiquitous high-quality communication of all types of information, including voice, data, fax, video, e-mail at, preferably, high data rates with high availability and reliability, with ensured privacy, using low-cost, low-powered devices, and at modest subscription or service fees.

It is not possible to make a full argument in this

space that wireline-based telephone and data services are not cost efficient for many parts of the United States. Low subscription rates for telephone service in areas with relative low population density are a result of cross-subsidization brought about by the principle of Universal Service which sets the rules for the telephone companies. Studies have indicated that a comparison, on a full cost basis, of the wireless service with the wireline service for those areas gives the advantage to the wireless. Many are convinced that in low population-density areas where the new communication infrastructure needs to be built, wireless systems make more economic sense.

Wireless systems provide another advantage: rapid incremental deployment. One can start with a limited capacity system. Increasing the capacity of a cellular system can be achieved by decreasing the cell size and installing new base stations. The capacity of a wireline system is in an important way determined by the capacity of the wiring which is difficult to replace. Even in those densely populated places where wireline systems are available and functioning well, and where high speed data transmission application might warrant the use of optical fiber to the neighborhood, if not to the curb, demand for wireless services is increasing at an astonishing rate. People want untethered communication capabilities.

Personal Communication Network services cover a range of wireless capabilities including cordless phones, microcellular systems for high density, low mobility environments, and regular cellular systems (now second generation, digital systems) which provide the high-tier services for high mobility, medium density environments. Providing world-wide coverage for these Personal Communication Network capabilities requires scaling up the cell size to a point where satellites are required to act as the base stations. This needs to be done in a way that the user can keep using the small handheld terminals with a substantial battery life as are now used for the microcellular or the regular cellular service. Finally, Personal Communication Network services typically require limited bandwidth. The satellite systems being developed to support PCN are also narrowband.

3   The Internet

It is hardly necessary to introduce the world wide system of interconnected networks which goes under the name of the Internet. The Internet is on an amazing exponential growth path, both in terms of the number of users and in terms of traffic. The traffic is growing at a higher rate than the number of users. Part of the reason for this is the popularity of the World Wide Web hypertext system. The documents being accessed via the World Wide Web are typically multimedia. The bandwidth requirements for this type of access

far exceed those of the more traditional uses of the Internet, e.g., remote terminal sessions and electronic mail. Newer Internet applications, among them Internet telephony and video conferencing, will drive up further the communication link capacity requirements. Essentially, then, from a communications standpoint the Internet needs to be a broadband network with bandwidth on demand and quality-of-service guarantees. Optical fibers come to mind first when thinking about the technology to use to build the infrastructure for this broadband network.

But challenges are coming from the broadband satellite systems now on the drawing board. As we will see, these systems are also considerably more complex than those under development for the narrowband service. Using a system of packet switching satellites to provide the Internet service has distinct advantages over using a wireline system. First, depending on the chosen constellation, world-wide or near-world-wide coverage is basically automatic. Any single point on the earth, even if there is no infrastructure on the ground within hundred of miles around it, can be accommodated. Also, capacity can be re-allocated when general usage pattern change or during unusual circumstances, without moving infrastructure around. Further, proponents of satellite systems are strongly convinced of the cost advantages of satellites in providing bandwidth. The labor cost of replacing wireline media are well recognized.

As we shall see below, depending on the type of satellite system involved signal propagation delay or latency can be substantial and pose a limitation on the kind of communication applications that can be supported. Also, when using reliable protocols, which require that data be kept in the sender's buffer until acknowledged by the receiver, the effective transmission rate may be constant even if we greatly increase the channel capacity. If the buffer size is B and the round trip latency is T, the effective transmission rate cannot be larger than B/T. One approach to mitigate this problem is to develop special protocols for use by the satellite systems. But, introducing special protocols for satellite systems limits the possibility of integrating them seamlessly into existing wireline broadband networks such as the Internet. Another alternative is to develop satellite broadband systems using satellites with low orbits for which propagation delays are low. To obtain the needed bandwidth, satellite systems being designed for broadband communications, use the Ka-band (roughly 18-31 GHz). This carries with it strong signal attenuation by rain. This problem is partly alleviated by having a sufficiently large number of satellites in the system, so that alternative paths exist to each service point.

4   Satellite Communication

The International Telecommunications Union (ITU) is responsible for defining the satellite services and assigning frequencies. There are some 17 different services defined by the ITU. These include: Fixed Satellite Service; Broadcast Satellite Services (otherwise known as Direct Broadcast Satellite Service (DBS)); Mobile Satellite Service and Radio Navigation Satellite Service. Satellite services which encompass multiple countries are typically established with the cooperation of local partners, such as PTTs or Public Switched Telephone operators, since this is frequently the easiest way to obtain the necessary licenses for the importation of equipment or the establishment of service. It is in this vein that INTELSAT was formed to provide fixed satellite services, and that it grew to where it has 130 member countries, and that INMARSAT was created to provide international mobile satellite services, reaching a membership of 75 [9].

The earliest communication satellites acted simply as reflectors: a sender aims the signal to the satellite which reflect it back to the recipient of the signal. Later came satellites which would provide signal amplification, as well frequency shifting, using different bands for the uplink and downlink. Current satellites can do actual processing of the signals, perform switching and use different beams to steer the signal in different directions on the downlink.

Satellites are naturally classified according to the size of their orbits. A satellite's orbit size or altitude above the earth determines a number of parameters which have important implications for its suitability for different purposes. One of them is propagation delay, or more generally, latency, and, as has already been alluded to, refers to the time it takes for a signal to travel from a sender on the earth up to the satellite and back down to a receiver on the earth. Most earth satellites fall in three categories: geostationary, medium earth orbit and low earth orbit [10].

4.1   Geosynchronous (GEO) Satellites

Geosynchronous satellites have orbits which lie in the plane of the equator, have a radius of about 42000 km and have the property that their periods are equal to the length of a day on earth. Consequently, a geosynchronous satellite, as observed by a fixed observer on earth, remains in the same apparent position in the sky. This means that earth-based antennas do not require tracking mechanisms: once pointed at the satellite, no further adjustment is required. Another advantage of GEO satellites is that, because of there great distance from the earth, only three of them are needed to cover the entire surface of the earth. These advantages of the GEO satellites have caused the available spots to be filled. Avoiding interference between satellites requires a minimum angle of separation of about 2 degrees between satellites which use the same frequency bands.

There are a number of disadvantages associated with GEO satellites. Because of the very long distance between earth terminals and the satellite, signal attenuation is very great. This necessitates powerful transmitters and a high gain (large diameter) antennas to be used and makes it impossible to build lightweight terminals or phones. The propagation delay for GEO satellites is of the order of a quarter second (one hop). To get an acknowledgement back from the receiver takes again the same time. This makes the use of GEO satellites for real-time applications such as telephony, video conferencing, as well as for other fast-response computer applications far less than ideal. The impact of latency on the performance of reliable protocols was discussed above in Section 3.

The "footprints" of the GEO satellites are large and as such much of the energy is wasted over oceans or sparsely populated areas. Even when multiple, narrower beams are used, it is difficult to focus the energy narrowly and to reuse the same frequencies for different areas on the surface of the earth. Ideally, to minimize signal attenuation by the atmosphere the angle of elevation of the satellite should be 90 degrees. While this is the case for places along the equator, it is far from that for the most populated regions of Europe and North America. While a number of GEO satellite systems which provide PCN-like services exists, among them INMARSAT-3, the newer projects for global Personal Communication Networks do not involve GEO satellites.

GEO satellites, because of their wide coverage, are particularly well suited for broadcasting. In the context of the Internet they are being used to implement the "push technology". Under this concept large numbers of Web pages can be downloaded to many PCs at the same time. As long as enough subscribers agree on what information should be downloaded ahead of when they need it, a miniscule amount of bandwidth per user is needed for the large amount of information that each one receives. Basically the same type of receivers and antennas used for direct broadcast satellite television can be used for this type of World Wide Web application.

4.2   Low Earth Orbit (LEO) Satellites

LEO satellites have orbits with an altitude of anywhere between 200 and 1500 km above the earth, and as such move around inside the inner Van Allen belt. Their periods are around 100 minutes and they stay in view of a fixed observer for about ten minutes depending on actual altitude.

LEO satellites have considerable advantages over GEOs in certain respects. The low altitude of the

LEO satellite means that the propagation delay is low and that it is suited for use in real-time applications. The low altitude also means that signal attenuation is much less than in the case of the GEO satellites and that small, low power terminal devices with low gain antennas can be used. It further means that the signal beams have smaller footprints. This allows cellular systems to be build with smaller cells and higher frequency reuse.

As expected, not all the characteristics of LEO satellite systems are advantageous. Satellites in low earth orbits have a shorter life span than those in high orbits. For continued coverage in any one area a number of satellites are needed, and, of course, to have full coverage of the earth a large number of satellites are required. To keep that number reasonable one may need to accept satellite low elevation angles during communication. To establish communication between any two points on the earth a number of inter-satellite links may need to be used, increasing the propagation delay in correspondence with the distance (but no more so than for a wireline network). The need for inter-satellite links substantially increases the system complexity. In a LEO satellite system connections will often last longer than the time that a given satellite remains in view. This means that graceful call handover from satellite to satellite needs to be worked out.

Systems of LEO satellites are prime candidates for the establishment of global Personal Communication Network service. A LEO satellite system with inter-satellite links can provide a complete alternative to an earth bound wide area network such as the Internet.

4.3 Medium Earth Orbit (MEO) Satellites

MEO satellites as their names implies have orbits with dimensions of the order of 10,000 km, considerable larger than those of LEOs, but less than half of those of the GEO satellites. As expected, the characteristics of MEO systems fall somewhere between these of GEO and LEO systems, so that they have, to a lesser degree, the advantages and disadvantages of the both of these types of systems.

5   Commercial Satellite Systems for Personal Communication Networks

For Personal Communication Networks, as already mentioned, GEO satellites are not suited because of the power requirements and weight of the mobile terminals which they imply. Several PCN systems are being deployed with planned operation within the next few years. Most of these are LEO systems and are designed to be global systems supporting real-time applications involving both voice and data. We note that Personal Communication Networks can provide access to the Internet. In this context it would be a narrowband access to the Internet, and one not very well suited for extensive multimedia applications.

5.1   Iridium

The Iridium system [4,6] is developed by Motorola at a cost of 5 billion US dollars. The launch of the first five satellites for this system took place on May 5, 1997. The system is expected to be operational before the end of 1998.

The system is made up of a constellation of 66 satellites having circular orbits (altitude 780 km), arranged in 6 polar orbital polar planes, each having 11 satellites. Collectively the satellites will provide more than 3000 beams of which only about 2000 will be active at any point in time. The other ones will be switched off as they pass over the poles where the beams of the satellites in adjacent orbital planes overlap. Iridium employs a combination of time division multiple access (TDMA) and frequency division multiple access (FDMA) The spectrum used by Iridium is the L- band (at 1.6 GHz) for user communications, and the Ka-band for down-link and up-link communications (at 19 GHz and 29 Ghz, respectively) to and from earth stations.

Earth stations are used to provide connections to the public switched telephone network. Inter- satellite links make it possible to hand over calls between satellites. Normally a call will not be handed over between satellites, but will be routed via the nearest earth station, but the inter-satellite link makes it possible to bypass the public switched telephone network altogether. The mobile terminals (handsets) will be dual-mode, allowing a customer to use the local cellular system where available, and the satellite system where no ground-based system exists. The call processing methods are fashioned after those of the 2nd generation cellular European phone system GSM. Data transmission service at 2400 bits per second is also provided.

When first conceived, the Iridium project was intended to stress service to areas of the world with less-developed telephone markets and had a target subscriber population of about 1 million. At this time, the primary market for the service is professional travelers with a subscriber population of 3 million.

5.2   Globalstar

The Globalstar system [3] is scheduled to launch its first satellites in the first half of 1998. It is being built at a total cost of about 2.5 billion US dollars. The system will start with an initial number of 32 satellites. It is expected to be completed by the middle of

1999, at which time it will consist of a constellation of 48 satellites. The 48 LEO satellites will be deployed in 8 orbital planes. The orbits will be circular at an altitude of 1410 km. For any point with a latitude between 70 degrees north and 70 degrees south there will be at least one satellite in view. There will be double or better coverage between the latitudes of 25 and 50 degrees, either north or south.

Globalstar created a consortium of commercial telecommunications providers which will make the service available around the world in cooperation with the local regulatory authorities. All calls will go through the local land-based network, unless non-existent, and therefore will not by-pass national telephone systems and local control.

Globalstar uses a band around 1.6 GHz for the user to satellite link, a bound around 2.5 GHz band for the satellite to user link. Gateway to satellite and satellite to gateway links use bands around 5 Ghz and 7 Ghz respectively. Globalstar uses Code division multiple access (CDMA). The CDMA technique allows, besides efficient spectrum usage, a user terminal to be communicating with every satellite in its view and combine the signals received from them to improve the quality of the reception. Globalstar supports data transmission from user terminals at rates up to 9.6 kilobits per second. The repeaters on the satellites in Globalstar are of the "bent-pipe" variety, meaning that there is no switching or processing. They simply shift the signals from the uplink frequency to the downlink frequency. There is no communication between satellites, i.e. there are no inter-satellite links.

Globalstar aims to serve customers in two distinct categories:

1. customers who derive great value from ubiquitous communication: international business travelers, general aviation, long distance truck operators and the like,

2. countries with under-developed telephone systems for which Globalstar offers stationary telephones.

5.3   Odyssey

The Odyssey satellite system [7] when fully deployed will consist of 12 MEO satellites, arrayed in 3 orbital planes at an elevation of 10,350 km, in between the inner and the outer Van Allen Belt. The periods of their circular orbits will be 6 hours. The first satellites are scheduled to be launched near the end of 1998. The system is intended to be completed by the end of 1999.

Special to this system is that the satellites are steerable to keep their beams directed to a particular region of the earth. The beams create 37 cells with a diameter of 1,100 km. The amount of power going to each cell is controllable. These capabilities allow the system to focus the resources on where the customers are

as opposed to, e.g., on essentially empty oceans. Usually, customers will have two satellites in view. The user terminal, in this case, will communicate via only one satellite while monitoring the other to help decide when to switch satellites when the first one is about to disappear from view.

CDMA is used as the multiple access method. The same frequencies are used for the user link as are used by Globalstar, which also uses CDMA. For the earth stations, a band around 19 GHz and a band around 29 GHz are used for the downlink and uplink respectively. The repeater type is again of the "bent-pipe" variety and, thus, no inter-satellite links are provided in this system. As in the case of Globalstar the Odyssey service will include data communications at rates up to 9.6 kilobits per second.

6   Commercial Broadband (Internet) Satellite Systems

The negative effect of latency on the operation of satellite communication means that broadband satellite system designs favor LEO satellites. GEO satellites find their place in hybrid systems where their broadcasting capabilities are used for services where the information flow is highly asymmetric. Below we present an overview of two commercial systems under development: one, a pure LEO satellite broadband system, and the other a hybrid satellite broadband system.

6.1   Teledesic

The Teledesic satellite system [5,11] is a broadband communication system being designed by the Teledesic Corporation. Its estimated cost is $9 billion dollars. The first launch is planned for the year 2001 and the system is to become operational in the year 2002. In its initial configuration, it will consists of 280 satellites in 14 orbital planes. These LEO satellites would be on circular orbits at an altitude of less than 1400 km. The satellites perform on-board processing and switching. The inter-satellite links operate at 155.52 Megabits per second (in the initial design). The entire system is a packet switching network using adaptive routing.

The satellites will use Ka-band frequencies for communicating with the earth terminals: the 28.6-29.1 GHz band for uplinks and the 18.8-19.3 GHz band for the downlinks. As mentioned above, Ka-band radio signals are subject to strong attenuation by rain, especially when the raindrops are as large as they are in tropical rain storms. If a satellite is present under a high elevation angle, this attenuation may be kept in check. With the initial constellation there will always be a satellite in view under an angle of 40 degrees or more - from any place on earth.

A combination of multiple access methods are used.

Within a cell the uplink capacity allocation is done using Multi-Frequency Time Division Multiple Access (MF-TDMA), for the downlink Asynchronous Time Division Multiplexing Access (ATDMA). User terminals will typically use uplinks at data rates of 2 Megabits per second and downlinks at rates of 64 Megabits per second, but connections at 64 Megabits per second in both directions will also be possible. In any cell of 100 km radius the network will provide a capacity of 500 Megabits per second. Globally, Teledesic envisions supporting millions of simultaneous users.

Teledisc is conceived as a broadband network which intends to provide data rates, error rates and latencies comparable to those of optical fiber networks. With these parameters assured, the system shoud be capable to run all the various protocols designed for wireline networks without modification. This way the system can be a seamless extension to the existing networks either by adding extra capacity where the wireline networks have failed or are over-stressed, or by supplying services in geographical areas not reached by the wireline networks. Personal Communication Network services may be integrated in the Teledisc system but are not very important in its conception.

The system definitely fits in the Internet. Teledisc has advertised its system as "Internet-in-the- Sky". It can stand on its own or supplement the existing Internet by extending the geographic range and the capacity. Like the wireline Internet in the future it will support real-time applications, bandwidth on demand and quality-of-service guarantees and support the same user protocols as the wireline Internet.

6.2   Celestri

The Celestri system [2] is being designed by Motorola. It will be a hybrid system consisting of a LEO satellite system and a GEO satellite system. The plan is to have it operational in the year 2003. The LEO system will consist of a constellation of 63 satellites in 7 orbital planes. The orbital planes form an angle of 48 degrees with the equatorial plane. The satellites will have circular orbits with altitudes of 1400 km. Transmission on the uplinks and downlinks will use the Ka band: uplinks use bands 28.6 GHz-29.1 GHz and 29.5-35.0 GHz, downlinks use bands 18.8-19.3 GHz and 19.7-20.2 GHz. The number of user uplinks per satellite will be 432 and the number of downlinks will be 260. The satellites will have processing and switching capabilities. Inter-satellite links will use lasers and transmit data at 4 Gigabits per second. Collectively the satellites in the LEO system will be able to move an amount of data on the order of 80 Gigabits per second between various user terminal devices. The earth terminals will operate at data rates anywhere from 2.048 Megabit per second (European E-1 type service rates) up to 155.2 Megabit per second.

The GEO system is planned to be made up of 9 satellites, jointly providing coverage for the populated areas of the world. These satellites will have 4 wide area beams and 84 spots beams with a downlink capacity of 2.8 Gigabits per second. Since it is to be a hybrid, coordinated system and not two separate systems going under one name, it will enable applications which both require the broadcasting of large amounts of data to many users and the collection of limited response from those same users. For instance, interactive television polls could be realized using the system.

In general terms, the system is intended to provide the broadest range of communication services, including broadcasting, multimedia, real-time data, etc. to a wide variety of customers including broadcasters, large corporations, Internet operators, and even individual customers. Even though Personal Communication Network services could be accommodated through this system, it is not its intended application area. The Celestri system focuses on fixed, broadband, as opposed to mobile, narrowband terminals.

In terms of the Internet applicability, the LEO portion of this system, just like the Teledesic system, could be used to provide expansion and back-up of the Internet infrastructure. The GEO portion of the system could readily be employed to implement "push technology" for the World Wide Web. In combination the GEO and LEO systems can be used to advantage for any distributed Internet application where the broadcasting capability of the GEO satellites cuts down in a substantial way the traffic that would otherwise be generated by sending identical information to very large number of users.

7   Conclusion

In this paper we have looked at the possibilities of providing two types of services using communication satellites: Personal Communication Networks and broadband Internet services. Insofar as they have been provided in the past, it has primarily been achieved using wireline systems.

We defined the concept of Personal Communication Networks and explained why wireless systems are best suited to satisfy their requirements and argued the need to extend current wireless systems to encompass the globe. We motivated the idea that the future Internet is to be a broadband network with bandwidth on demand and quality-of-service guarantees.

After taking a look at the principal characteristics of geostationary, low earth orbit, and medium earth orbit satellites and satellite systems we reviewed some key commercial systems which are currently being deployed or are in an advanced design stage, and will provide one or the other of these two types of services. Notwithstanding the larger number of satellites required and the attendant greater complexity of LEO

systems, most of the current designs favor the LEO approach for both these types of services.

While any individual system is a risky enterprise, there is little doubt that the concepts they embody will ultimately succeed in some form and bring one more paradigm shift in personal communication and computing.

References

[1] Calhoun G. (1988) *Digital Cellular Radio*, Artech House, Norwood, MA.

[2] Celestri, [http://www.mot.com].

[3] Globalstar, [http://www.globalstar.com].

[4] Hardy Q. (1996) Satellite-Data Firm Hopes to Prove It's not Pie in the Sky, Wall Street Journal, December 20.

[5] Hardy Q. (1997) Iridium Phone Project Maps an Upscale Orbit, Wall Street Journal, January 10.

[6] Iridium, [http://www.iridium.com].

[7] Odyssey, [http://www.trw.com/seq/sats/-ODY.html].

[8] Paetsch M. (1998) *Mobile Communications in the U.S. and Europe*, Artech House, Norwood, MA.

[9] Pelton J. N. (1995) *Wireless and Satellite Communications*, Prentice Hall, Upper Saddle River, NJ, p. 123-160.

[10] Tanenbaum A. (1996) *Computer Networks*, Prentice Hall, Upper Saddle River, NJ, p. 163-170.

[11] Teledesic, [http://www.teledesic.com].

Patterns in a Hopfield Linear Associator as Autocorrelatory Simultaneous Byzantine Agreement

Paule Ecimovic
University of Ljubljana, Department of Philosophy, Aškerčeva 2, Ljubljana, Slovenia
Phone: +386 61 176 9200 int. 386
E-mail: `ecimovic@ctklj.ctk.si`

In the first part, the Byzantine Agreement problem in parallel distributed processing is formulated for generalized, completely-interconnected networks of interacting processors. An overview of the main cases of this problem are presented in brief. Among standard optimal algorithms for reaching Simultaneous Byzantine Agreement, only the two-phase commit protocol is set out in any detail. In the second part, the process of pattern formation in Hopfield linear associators, realized as single-layer neural networks with Hebbian weight adjustment rules, is discussed. The main result of the paper is then presented, according to which pattern formation in Hopfield linear associators is a solution to a form of Simultaneous Byzantine Agreement. In conclusion, it is argued that such associative memory solutions to interactive consistency problems in generalized transaction processing systems may finally prove viable, despite decades of neglect due to inavailability or prohibitive expense of sufficient processing power for their large-scale implementation.

1 Introduction

The transaction processing problem[1] of achieving and/or maintaining interactive consistency, as typified by the Byzantine Agreement (BA) scheme (Pease *et al.* 1980), (Lamport *et al.* 1982), formulated below, is exemplary among workable approaches to the design and implementation of fault-tolerant distributed protocols in parallel distributed processing (PDP) (transaction) systems. Wide-spread commercial applications, especially in banking, depend critically on guaranteeing a minimally-sufficient reliability of correct execution of distributed protocols in the presence of faults, be they faulty connections or faulty (or ill-synchronized) processors, or combinations of both (Wang 1995: p.420). In particular, achieving interactive consistency among inter-connected and distributed processors in the presence of faults is an important problem to which standard round-optimized solutions exist, such as the Dwork-Moses protocol or the two-phase commit protocol. The time might well have come, however, to (re-)consider the advantages of associative strategies typified by Hopfield's linear associator (LA) with a Hebbian association rule, most

commonly realized as a fully-connected, single-layer neural network, where the connection strengths are determined by a form of Hebb's learning rule.[2] Before expanding on the close analogy between BA and associative pattern formation in the Hopfield LA, let us consider the BA scheme, its most prevalent instances, and some round-optimized protocols for reaching various instances of BA.

2 Byzantine Agreement (BA)

2.1 Fundamental Definitions

The fundamental ingredients of problem of reaching Byzantine agreement are the following. We take a processor to be any suitable Turing machine capable of carrying out some effective elementary instruction-set. Theoretically, this can be construed as a universal Tur-

---

[1]The author is currently funded by the Slovene Science Foundation, while pursuing a master's degree in logic at the College of Philosophy, at the University of Ljubljana, where he is also engaged in research.

[2]We have deliberately avoided establishing an "understood" transition to neural networks *per se* in order to facilitate realizations of the Hopfield LA scheme in other physical systems, which might not behave "neurally" at all, in any currently-simulated sense. We feel that this gives both neural information processing researchers an opportunity to reappraise the (in)adequacy of their neuron models from an information processing point of view as well as the neuro-biological community a breather from computational neuron-modelling strategies, which have recently been shown to miss a vast array of interaction detail. See: (Koch 1997) for hints.

ing machine. Practically, this can be anything as simple as an "intelligent" bistable switch enhanced with some automated processing logic, e.g., a McCullough and Pitts binary neuron, or something as complex as a human operator sitting at a computer terminal connected to a local area network. Some interconnected network of such processors is required across which to distribute and on which to execute a protocol thus distributed. By *distributed protocol*, we are to understand any non-contradictory set of instructions P which can be divided into n parts P_i such that

$$P = \bigcup_{i=1}^{n} \{P_i\}$$

each part of which is assigned to a given processor in the course of a task designation phase of (pre-)processing of the protocol. Atop this layer, we must have a meta-layer for regulating and monitoring inter-processor communication in such a way as to achieve *interactive consistency*, which means in this context that none of the processors contradicts any action(s) of other processors on their data or on its own data or of itself on its own data. Figuratively speaking, if one were to set a team of people to sweep a room, there would be a supervisor to ensure that no member of the team throws dust on an area just swept by another member or by the member him- or her- self, thereby guarding against "stupid mistakes"(Novak,1993:p.27). We also assume a *global clock* (generator of regular events) relative to which all inter-processor communication in the PDP network is synchronized and measured. A *round of computation* is defined as the time interval (number of regular events) generated by the global clock required for all the parts P_i of protocol P to be executed by the corresponding processors p_i *once.* [3] Informally, a *run* is the entire state transition history of all the states in which all the processors were in the course of all rounds of computation of a given protocol P from some arbitrary starting tick of the global clock to some other arbitrary tick, i.e., in some time interval as measure on the global clock.[4] For practical reasons, we often normalize the clock ticks to coincide with rounds of computation of a given protocol, making the obvious execution uniformity assumptions. By *crash failure* is meant the state of a PDP system in which execution of a protocol from a certain round onward generates results incompatible with any admissible run of the given protocol. [5] The class of parallel-distributed processing (PDP) systems

to be considered for the rest of the paper is defined by all PDP systems, which satisfy the following five conditions (Wang 1995: p.420): For natural numbers k and n, such that $n \geq k + 2$:

PDP1 There are n processors, at most k of which are faulty with respect to a given semantics, whether functional, operational, or declarative (elaborated below), without incurring crash-failure of the system in executing a given distributed protocol P

PDP2 the processors can communicate directly with each other through message exchange in a fully-connected network (FCN)

PDP3 the message sender is always identifiable by the receiver

PDP4 an arbitrary set of processors are chosen as sources and their initial value v_s is broadcasted to other processors and to themselves at the start of parallely-distributed execution of P

PDP5 The only faulty components considered are processors.

PDP4 is a significant generalization from the PDP specifications presented in (Wang,1995: p.20), since it allows for more than one processor to carry the initial state which is to be distributed to and agreed upon by the entire PDP network. This is critical for our LA realization, because it admits the case of an "initial configuration", see the next section, being distributed initially across the entire network. PDP1, *mutatis mutandis*, can be taken as a general definition of what it means for some protocol P to be *k-resilient.* In other words, a protocol P is k-resilient if, and only if, for $k \leq n+2$, (at most) k processors can fail and P still executes and terminates correctly. Thus, k-resilient distributed protocols are prime examples of fault-tolerant design. The following are ways in which a processor can be *faulty*. A processor p_i is termed *faulty* if it satisfies either or both of the following conditions:

– it fails to send *appropriate messages (defined either by a monitoring meta-level algorithm or at the object protocol level and just monitored "from above")*

– it fails to send otherwise appropriate messages at a time on the global clock foreseen for it in a given protocol.

A processor which is not faulty is termed *healthy.*

---

[3]Henceforth, processors will be denoted by lower case subscripted p's, whereas the part of a given protocol P assigned to each by upper case subscripted P's.

[4]Formally, this can be defined in terms of the modal logic S_5, where it is possible to define the set of all possible state transitions and quantify over them with possibility and necessity operators. (Halpern 1986)

[5]Crash failure could be defined more "severely" and absolutely as a state of a PDP system after which the system ceases

to function. This definition is vague and thus prone to all sorts of error and might better be avoided.

2.2 The BA Scheme and its major variants

An n-processor PDP system, satisfying PDP1-5 of which at most k processors are or prove faulty, in the above sense, reach **Byzantine Agreement** if, and only if, (Wang,1995: p.20):

BA1 Every healthy processor computes the common value v which is used to determine the agreement

BA2 If the source, in the sense of PDP4, is healthy, then the common value should be the source's initial value v_s.

There are two major variants to the above general scheme, which we call BA. The one which we will consider is called **Simultaneous Byzantine Agreement** and is obtained from the above scheme by adding the following condition:

SBA3 All healthy processors reach agreement in the sense of BA1 & BA2 *in the same round of computation.*

We mention the other major variant in passing, namely, **Eventual Byzantine Agreement**, which is obtained by adding the following requirement to BA1,BA2:

EBA3 Each healthy processor will reach agreement with all other healthy processors, in the sense of BA1 & BA2 *in some round of computation, if the run lasts long enough with respect to the global clock.*

2.3 Round-optimized BA Solution Algorithms

The solution algorithms for BA mentioned or presented here impose more stringent restrictions on the number of faulty processors than we imposed in the framework above, in which solutions, let alone optimal solutions, are not guaranteed. In this respect, the above framework can be taken heuristically a means of exploring the problem space, rather than the solution space. In their original treatment, Pease, Shostak, and Lamport see (Pease *et al.* 1980) solved the so-called Byzantine Agreement problem for a fully-connected network. The formulation of the problem axiomatized as BA1 and BA2 above along with its variants are derived from (Pease *et al.* 1980), (Lamport 1982), (Halpern 1986), and (Wang 1995). For a fully-connected system of n units, BA is reachable if, and only if, $n \geq 3k+1$. This formulation leaves malicious units anonymous (unidentified) in the sense of (Novak 1993: p.22). For n=4 and k=1, two "phases" (rounds of computation) suffice to reach BA. In our notation, the situation is as follows, denoting messages sent by processor p_i by m_i. See (Novak 1993) for a diagram

and re-label accordingly. Here the specification of P and its partitioning is unimportant, since it applies to most all parallelly-distributable protocols, with some solvability and other application-specific restrictions. Processors p_1, p_2, and p_3 send their messages, m_1, m_2, and m_3 to the other three processors respectively, whereas processor p_4 sends m_4 only to p_2 and p_3, but not necessarily to p_1. Instead, p_4 sends p_1 some message X. Assuming that for $i \in \{1,2,3\}$, p_i is fault-free and that p_4 is faulty in the above sense as well as being "maliciously" faulty, i.e., sending some message X, and claiming it sent the expected message, m_4, with the claim not being necessarily true (Novak 1993). Thus, the algorithm for reaching BA in this case proceeds in the following two phases:

TPS1 Each processor p_i sends its message to all the other processors

TPS2 Verification of message correctness.

Thus BA is reached, according to the above specifications, and p_1 can claim to have received message X from p_4 as a result of the TPS2 phase, whereas p_4, being "malicious" can claim that it sent m_4. Thus we have reached "Byzantine" agreement, agreeing, within reason, to disagree, while still getting the job done. Comparison TPS1 and TPS2 with the **two-phase commit** protocol for distributed databases (McFadden 1994: p.481) reveals a similar pattern, with some modification relating to the application domain, etc. Suffice it for our purposes to interpret "site" as "processor" and "commit" as "agree on value a common v as per BA1" For explication of the broadcast variant of BA, see (Wang,1995).

TPC1 A message is broadcast to every participating site, asking whether that site is willing to commit to its portion of the transaction at that site. Each site returns an "OK" or "not OK" message.

TPC2 The originating site collects messages from all sites. If all are "OK," it broadcasts a message to all sites to commit the transaction (come to agreement). If one or more responses are "not OK," it broadcasts a message to all sites to abort the transaction.

Obviously, this is a "no non-sense" (fault intolerant) specification which could be generalized to admit one or more failures to respond "OK" in which case it would amount to a direct solution to BA. However, non-sense at the level of automated teller machines savings/checking account transactions and corresponding account balance updates is not taken lightly by the customer on the street who is likely to complain angrily at the slightest "irregularity", let alone fault. At this level, fault-tolerance and k-resilience are not to affect the end user, but rather only the processing system and its administrators.

3   Hopfield's Linear Associator (LA) and its Convergence Properties

In this section, we define Hopfield's linear associator (LA) in terms of discrete-time dynamical systems (effectively, finite state automata) and weighted graphs, thus avoiding this systems usual interpretation as a neural network. The reason for this will become apparent in the final two sections of this paper.

3.1   Specification of the Hopfield Linear Associator

The topology and architecture of Hopfield's LA (HLA) are specified as follows (Bruck 1990:p.247):

HLA1 A Hopfield LA is a discrete-time dynamical system represented by a weighted graph

HLA2 Each edge of the graph is assigned a weight and each node a threshold value.

HLA3 The weighted graph is fully connected.

The *order* of the HLA is defined to be the number of nodes in the corresponding graph. Thus, if N is an HLA of order n, written henceforth N∈HLA(n), then, according to HLA1-3, the ordered pair (W,T) determines the topology and architecture N uniquely to within a renaming, where W and T are defined as follows. For natural number n:

- W is an n x n matrix, the elements w_{ij} of which represent the weight assigned to the edge joining the i-th and j-th nodes.

- T is a vector of dimension n, the components t_i of which denote the threshold assigned to the i-th node.

The processing element represented by any node can be in one of two possible states, either -1 or +1, and the state is denoted by $s_i(t)$, where t is a natural number to be interpreted as a certain discrete number of ticks of a clock analogous to the global clock defined above. The n-dimensional vector $s(t)$ of all elements $s_i(t)$ represents the state of N(n). The vector s is called the *state vector* of the HLA N(n). Thus, the order of N is determined by the dimensionality of its state vector. Furthermore, the state vector represents the states of all individual processing elements ("processors", for short) represented by the nodes of the graph. Thus, the state vector $s(t)$ is defined by the equation

$$s(t) := (s_1, s_2, \ldots, s_n). \qquad (1)$$

All the set of all possible state vectors of a suitably-chosen set of individual processor states s_i, where i

varies from 1 to n by ones and n is some finite natural number, forms a *state space*. The evolution equation of the dynamical system, i.e., the equations governing the transition of the system from one state $s(t_0)$ to the next state $s(t_0 + 1)$ are given componentwise as follows:

$$s_i(t_0 + 1) = sgn(H_i(t_0)), \qquad (2)$$

where the function $sgn(x)$ is the sign of the number x defined as +1, if $x \geq 0$ and -1, otherwise, and

$$H_i(t_0) = \sum_{j=1}^{n} w_{j,i} s_j(t_0) - \mathbf{t_j}. \qquad (3)$$

Accordingly, each processor at a given node is a linear threshold element, which adds its input signals from all other elements, at all other nodes in the system. Thus, each linear threshold element acts here as an "adder". To complete the specification of HLA, we must state a Hebbian weight adjustment rule and an energy function as follows (Hopfield 1982: p.2556).

$$\Delta w_{j,i} = [s_i(t)s_j(t)]_{average}, \qquad (4)$$

where the average is calculated over past history, and

$$E = -\frac{1}{2} \sum_{i \neq j} \sum w_{j,i} s_i s_j \qquad (5)$$

and the change in energy ΔE due to a change in state ΔE due to change in the state of an individual processor Δs_i is given by

$$\Delta E = -\Delta s_i \sum_{j \neq i} w_{j,i} s_j. \qquad (6)$$

Apart from particular initial conditions, which depend on the specific problem instance in question, this completes our specification of HLA.

3.2   Convergence Properties of the HLA

The feature property of the HLA for the purposes of reaching BA and, more specifcally, SBA, is that since its state space is finite, the HLA dynamical system will *always* coverge to either a stable state/cycles in state space (Bruck 1990). A specific value of the state vector $s(t)$ at a given time t is called a *configuration*. Stable states (or for a given t, configurations) of the HLA are called *patterns* (or *pattern configurations*, respectively). Thus, *the process of n processors reaching SBA reduces to the convergence of the HLA to a pattern or pattern configuration.*

4   Reaching Simultaneous Byzantine Agreement with Hopfield's LA

In this section, we state the main result of this paper, which in the terminology introduced above, may be stated as **SBA can be represented by a pattern in an HLA.** This implies that the process of reaching SBA can be represented by convergence of an HLA to a stable configuration, i.e., to a pattern. Now, we must just state some facts about the stability of states and configurations in order to indicate corresponding conditions for achieving SBA as well as flush out a key property of HLA's which corresponds to k-resilience and realizes fault tolerance.

4.1   Stability of Configurations

A necessary and sufficient condition for a state of an HLA to be stable is the following (Bruck 1990: p.247) (in vectorial form):

$$s(t) \; is \; stable \Leftrightarrow s(t) = sgn(Ws(t) - T), \qquad (7)$$

where W is an n x n matrix of weights and T is an n-dimensional vector of node thresholds. This is critical in determining patterns, since for a given time t, the state $s(t)$ becomes a configuration, which, if it is stable, is a pattern.

4.2   k-resilience and Fault Tolerance in HLA

As is evident from the HLA specification given above, the next state at time $t_0 + 1$ of an HLA is computed componentwise from its current state at some time t_0 applying equation (2) to each component $s_i(t)$ of $s(t)$. This yields the state of the HLA at time $t_0 + 1$, i.e., $s(t_0 + 1)$. One of the most prominent features of the HLA model which make it well suited to the implementation of fault-tolerant distributed protocols is that $s(t_0 + 1)$ can be computed from a *proper subset*[6] of all processors in the system (Bruck 1990). This is analogous to k-resilient protocols where, by definition (Kilmer 1995), at most k processors could fail, yet the protocol continues to execute and terminates correctly. Specifically, if S is the set of indices of processors from whose states the state of $N(n)$ is computed, then

$$cardinality(S) \leq n \qquad (8)$$

and

$$cardinality(complement(S)) = k, \qquad (9)$$

where n and k have the same meanings as in the PDP and SBA specifications in the previous sections.[7]

5   Conclusion

In this paper, we have demonstrated that Simultaneous Byzantine Agreement can be represented by a Hopfield linear associator. We have done so without interpreting Hopfield's system as a neural network, but rather as a discrete dynamical system. This allows for implementations in other physical systems, which have possibly greater computational power (as measured by the classes of problems which can be solved by means of such systems) than the current computational models and systems used to implement artificial neural networks. This will be the topic of a forthcoming paper.

Acknowledgements

---

[6]If set A is a *subset* of set B, then every element of A is an element of B. A set C is a *proper subset* of set B if B contains at least one element which is not contained in set C.

[7]The *cardinality* of a set is the number of elements contained in the set, and the *complement* of a given set is the set of all those elements which are not contained in the given set.

References

[1] Bruck J. (1990) On the Covergence Properties of the Hopfield Model *Proceedings IEEE*, Vol. 78, No. 10, Oct., p. 1579-1585.

[2] Chakrabarti B.K., Dutta A., & Sen P. (1996) *Quantum Ising Phases and Transitions in Transverse Ising Models*, Berlin: Springer-Verlag.

[3] Halpern J. (ed.) (1986) *Theoretical Aspects of Reasoning about Knowledge*, Proceedings of the 1986 Conference, Los Altos: Morgan Kaufmann.

[4] Kilmer W. L. (1995) Self-Repairing Processor Modules *IEEE TRANSACTIONS ON RELIABILITY*, Vol. 44, No. 2, June

[5] Koch C. (1997) Computation and the single neuron, *Nature*, V ol. 35, 16 January, p. 207-210.

[6] Lamport L., *et al* (1982) The Byzantine Generals Problem *ACM Trans. Programing Languages, Syst.*, Vol.4, No.3, pp. 382-401, July.

[7] McFadden F.R. & Hoffer, J.A. (1994) *Modern Database Management*, 4th ed., Redwood City: The Benjami/Cummings Publishing Company, Inc.

[8] Novak F. & Klavžar S. (1993) An Algorithm for Identification of Maliciously Faulty Units, *International Journal of Computer Mathematics*, Vol. 48, p.21-29.

[9] Pease M., Shostak R., and Lamport L. (1980) Reaching agreement in the presence of faults JACM, Vol.27, No.2, pp.228-234, April.

[10] Wang S.C., Chin Y.H., & Yan K.Q. (1995) *IEEE Transactions on Parallel and Distributed Systems*, Vol.6, No.4, April, p.420-427.

Data Fusion of Multisensor's Estimates

Yonggun Cho and Jin H. Kim
Department of Computer Science, Korea Advanced Institute
of Science and Technology,
Kusong-dong, Yusong-gu, Taejon, 305-701, Korea
Phone: +82 42 869 3557, Fax: +82 42 3510
E-mail: ygcho@ai.kaist.ac.kr
AND
Vladimir Ignat'evich Shin
Department of Mathematics, Institute of Informatics
Problems, Russian Academy of Sciences,
30/6, st. Vavilova, Moscow, 117900, GSP-1, Russia

The objective of data fusion is to combine elements of raw data from different sources into a single set of meaningful information that is of greater benifit than the sum of its contributing parts. In (Yonggun Cho et al. 1996) we developed data fusion methods for a class of linear and nonlinear continuous dynamic systems with multidimensional observation vector. In this paper we present a generalization of these methods to the fusion of dependent estimates and also to discrete stochastic systems determined by difference equation. The proposed fusion methods allow fully parallel processing of information and fit in with multisensor environment. Examples demonstrate the accuracy and efficiency of the proposed fusion methods.

1 Introduction

Most people would agree that artificial intelligence is a technological domain that encompasses a wide range of disciplines and techniques, and thereby finds application across a diversity of problems. In a sense, the data fusion domain has similar qualities, and the breadth of both fields makes their intersection.

In recent years, there has been growing interest to fuse multisensory data to increase the accuracy of estimation of dynamic systems (Ren C.Luo, Michael G.Kay 1989). This interest is motivated by the availability of different types of sensor which uses various characteristics of the optical, infrared and electromagnetic spectrums. In many situations targets are tracked by multisensors. The measurements used in the estimation process are assigned to a common target as a result of the association process. Once the targets represent the common target, then next problem is how to combine the corresponding estimates to get more accurate estimates. The well-known formula (Bar-Shalom and Leon Campo 1986) for the two-sensor track fusion was shown where the two estimation errors are correlated. But there is a need to generalize this formula for the multisensor environment so that we can fuse more than two dependent estimates.

This paper is the accommodation of the correlated estimation error to the theory of data fusion formula (Yonggun Cho., et al. 1996) and generalization of these methods to the discrete Kalman filter. So we can fuse any number of N correlated estimates via the use of a decomposition of observation vector. Usual Kalman filter (or extended Kalman filter) is replaced by local filters which reduces computational loading, achieves higher survivabilty, and is very suitable for a distributed sensor-level tracking. Examples show reduced computations and increased accuracy.

2 Problem Statement

Consider a kinematic model of a tracked target described by the differential equation:

$$\dot{x} = Fx + Gw, \quad x(t_0) = x_0, \ t \geq t_0, \qquad (1)$$

where $x = x(t) \in R^n$ is the target state vector of the system, $F = F(t)$ and $G = G(t)$ are $n \times n$ and $n \times r$ matrices, respectively. The stochastic process (input noise) $w = w(t) \in R^r$ is a Gaussian white noise with zero mean and covariance matrix

$$E[w(t)w(\tau)^T] = Q(t)\delta(t - \tau).$$

Here R^n is an n-dimensional Euclidian space, E denotes the expectation, superscript T represents the

transpose of a matrix or vector, and δ is the Dirac delta function. Assume that multisensor system is composed of N different types of sensors equipped with Kalman filters determined by the following equations:

$$y_k = H_k x + v_k, \quad k = 1, \ldots, N, \qquad (2)$$

where $y_k = y_k(t) \in R^{m_k}$ is the observation vector and $H_k = H_k(t)$ is the $m_k \times n$ matrix. The stochastic process(observation error) $v_k = v_k(t) \in R^{m_k}$ is a Gaussian white noise with zero mean and covariance matrix

$$E[v_k(t) v_k(\tau)^T] = R_k(t)\delta(t - \tau), \quad k = 1, \ldots, N.$$

We shall also assume that the measurement noises $v_1(t), \ldots, v_N(t)$ are independent of each other, so that $E[v_i(t) v_j(\tau)^T] = 0$ for all $i \neq j$ and $t, \tau \geq t_0$.

Note that N filtering errors are correlated, even though we assume that the measurement noises $v_1(t), \ldots, v_N(t)$ are independent of each other. Here it is required to estimate the state vector $x(t)$ by the minimum mean square error criterion on the basis of observations $\{y_k(\tau), \ t_0 \leq \tau \leq t, \ k = 1, \ldots, N\}$, taking account of correlations of filtering errors.

3 Fusion of the Correlated N Estimates

Decomposition of observation vector. We denote an optimal (in mean square sense) estimate of $x(t)$ based on observations $\{y_k(\tau), \ t_0 \leq \tau \leq t, \ k = 1, \ldots, N\}$ by $\hat{x}(t)$. In this case it takes the form:

$$\begin{aligned} \dot{\hat{x}} &= F\hat{x} + PH^T R^{-1}(y - H\hat{x}), \\ \hat{x}(t_0) &= m_0, \end{aligned} \qquad (3)$$

$$\begin{aligned} \dot{P} &= FP + PF^T - PH^T R^{-1} HP + GQG^T, \\ P(t_0) &= P_0, \end{aligned} \qquad (4)$$

where $y = [y_1^T \ \ldots \ y_N^T]^T \in R^m$ is the composite observation vector $(m = \sum_{k=1}^N m_k)$, $H = [H_1^T \ \ldots \ H_N^T]^T$ is the $m \times n$ matrix, $R = diag[R_1 \ \ldots \ R_N]$ is the $m \times m$ covariance matrix of m-dimensional composite white noise $v = [v_1^T \ \ldots \ v_N^T]^T$, and

$$P = E[(\hat{x} - x)(\hat{x} - x)^T]$$

is the $n \times n$ covariance matrix of filtering error $\tilde{x} = \hat{x} - x$.

According to (1) and (2), we have N different subsystems. Now we propose a new suboptimal filter (fusion method) taking account of the correlations of N estimates. We denote an estimate of vector x based on observations $\{y_k(\tau), \ t_0 \leq \tau \leq t\}$ by $\hat{x}_k$. To find the estimate $\hat{x}_k$ we use Kalman filter to the linear subsystem (1) and (2). We have

$$\begin{aligned} \dot{\hat{x}}_k &= F\hat{x}_k + P_k H_k^T R_k^{-1}(y_k - H_k \hat{x}_k), \\ \hat{x}_k(t_0) &= m_0, \quad . \end{aligned} \qquad (5)$$

$$\begin{aligned} \dot{P}_k &= FP_k + P_k F^T - P_k H_k^T R_k^{-1} H_k P_k + GQG^T, \\ P_k(t_0) &= P_0, \end{aligned} \qquad (6)$$

where

$$P_k = E[(\hat{x}_k - x)(\hat{x}_k - x)^T] \qquad (7)$$

is the $n \times n$ covariance matrix of filtering error $\tilde{x}_k = \hat{x}_k - x$.

Fusion of N-estimates of common process. If the common process hypothesis is accepted, then from (5) and (6) we have N filtering estimates $\hat{x}_1, \ldots, \hat{x}_N$ based on observations $\{y_1(\tau), \ t_0 \leq \tau \leq t\}, \ldots, \{y_N(\tau), \ t_0 \leq \tau \leq t\}$, respectively. Then new suboptimal estimate $\hat{x}^*(t)$ of the state vector $x(t)$ based on observations $\{y_k(\tau), \ t_0 \leq \tau \leq t, \ k = 1, \ldots, N\}$ is constructed from the estimates $\hat{x}_1, \ldots, \hat{x}_N$ by the following formula:

$$\hat{x}^*(t) = \sum_{k=1}^N c_k(t)\hat{x}_k(t), \quad \sum_{k=1}^N c_k = I, \qquad (8)$$

where $c_1, \ldots, c_N$ are known weighting coefficients and I is unit matrix.

The well-known formula (Y. Bar-Shalom and Leon Campo 1986) for the optimal combination of two($N = 2$) dependent estimates $\hat{x}_1$ with covariance P_1, $\hat{x}_2$ with covariance P_2 of x is

$$\hat{x}^* = \hat{x}_1 + (P_1 - P_{12})(P_1 + P_2 - P_{12} - P_{21})^{-1}(\hat{x}_2 - \hat{x}_1), \qquad (9)$$

where P_{12} and P_{21} are cross error covariances between two dependent filtering errors. This formula can be generalized to any number of dependent estimates $\hat{x}_1$ (with cross covariances $P_{11}, \ P_{12}, \ldots, P_{1N}$), $\ldots$, $\hat{x}_N$(cross covariances $P_{N1}, \ P_{N2} \ldots, P_{NN}$). Namely,

$$\hat{x}^*(t) = \sum_{k=1}^N c_k \hat{x}_k(t), \qquad (10)$$

where

$$\sum_{i=1}^{N-1} c_i(P_{ih} - P_{iN}) + c_N(P_{Nh} - P_{NN}) = 0,$$

$$(h = 1, \ldots, N - 1), \qquad \sum_{k=1}^N c_k = I \qquad (11)$$

represents the linear algebraic Eqs. for matrices $c_1, \ldots, c_N$.

This fusion process is described in Figure 1 and the proof of the Eqs. (11) is given in Appendix. Note that the equations (5) and (6) are separated for the various values of $k = 1, \ldots, N$. Therefore it could be solved in parallel and fit for the sensor level tracking system in the multisensor environment.
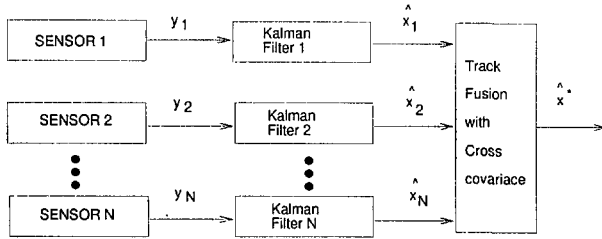
Figure 1: Multiple track fusion of N-estimates of common process

If we assume that the filtering errors $\tilde{x}_i(i = 1,\ldots,N)$ are independent (or uncorrelated), i.e., $P_{ij} = 0$ at $i \neq j$, then the Eqs. (11) take the simple form:

$$c_h P_{hh} - c_N P_{NN} = 0 \quad (h = 1,\ldots,N-1), \quad \sum_{k=1}^{N} c_k = I \tag{12}$$

or

$$c_h = c_N P_{NN} P_{hh}^{-1} \quad (h = 1,\ldots,N-1), \quad \sum_{k=1}^{N} c_k = I. \tag{13}$$

Substituting the formulae for $c_1,\ldots,c_{N-1}$ into Eq. $\sum_{k=1}^{N} c_k = I$, we have the exact formulae for coefficients c_h, $h = 1,\ldots,N$:

$$c_h = (\sum_{k=1}^{N} P_{kk}^{-1})^{-1} P_{hh}^{-1}, \quad h = 1,\ldots,N. \tag{14}$$

This results coincide with suboptimal filtering in (Yonggun Cho., et al. 1996). But the assumption that the filtering errors $\tilde{x}_i$ and $\tilde{x}_j$ are independent, i.e., $P_{ij} = 0$ at $i \neq j$, in general case is wrong. Therefore the formula (14) are only approximate formula. That means the coefficients c_h determined by formula (14) don't provide minimum of mean square error. However the formula(14) can be used in practice, especially in steady-state conditions, when the cross-covariance between errors $P_{ij} \approx 0$. The covariance matrix $P^*(t)$ of the filtering error can be calculated as shown in (Yonggun Cho., et al. 1996):

$$
\begin{aligned}
P^* &= E\left[\left(\sum_{i=1}^{N} c_i(\hat{x}_i - x)\right)\left(\sum_{j=1}^{N} c_j(\hat{x}_j - x)\right)^T\right] \\
&= \sum_{i,j=1}^{N} c_i E[(\hat{x}_i - x)(\hat{x}_j - x)^T]c_j^T \\
&= \sum_{i,j=1}^{N} c_i P_{ij} c_j^T, \tag{15}
\end{aligned}
$$

Note that at $i = j$ the covariance matrix P_{ij} coincides with covariance matrix P_i (7) determined by Riccati equation (6), i.e., $P_{ii} = P_i$. And,

$$
\begin{aligned}
\dot{P}_{ij} &= (F - P_i H_i^T R_i^{-1} H_i) P_{ij} \\
&+ P_{ij}(F - P_j H_j^T R_j^{-1} H_j)^T \\
&+ GQG^T \quad \text{for } i \neq j, \tag{16}
\end{aligned}
$$

$$
\begin{aligned}
\dot{P}_{ii} &= (F - P_i H_i^T R_i^{-1} H_i) P_{ii} \\
&+ P_{ii}(F - P_i H_i^T R_i^{-1} H_i)^T \\
&+ GQG^T + P_i H_i^T R_i^{-1} H_i P_i. \tag{17}
\end{aligned}
$$

Taking into account that $P_{ii} = P_i$, we can easily show that equation (17) coincides with equation (6).

4 Example of Dependent Estimates Fusion

4.1 Steady state filter with simple data

Let the dynamic system be described by scalar equation

$$\dot{x} = ax + w, \quad t \geq t_0.$$

Multiple sensor(observation system) involves two sensors. Signals y_1 and y_2 received by two different impulsers contain the state variable x and noises v_1 and v_2:

$$y_1 = b_1 x + v_1, \quad y_2 = b_2 x + v_2.$$

Here w, v_1 and v_2 are independent zero-mean scalar Gaussian white noises with variances Q, R_1 and R_2, respectively. Further, it is assumed that $a < 0, c_1, c_2, Q, R_1$ and R_2 are known constants, and initial value $x_0 \sim N(m_0, P_0)$. The optimal filtering estimate $\hat{x}(t)$ based on observations $\{y_k(\tau), \quad t_0 \leq \tau \leq t, \quad k = 1, 2\}$ is determined by the Kalman filter equations (3) and (4):

$$\dot{\hat{x}} = a\hat{x} + P\left[\frac{b_1}{R_1}(y_1 - b_1\hat{x}) + \frac{b_2}{R_2}(y_2 - b_2\hat{x})\right],$$

$$\hat{x}(t_0) = m_0, \tag{18}$$

$$\dot{P} = 2aP - P^2\left(\frac{b_1^2}{R_1} + \frac{b_2^2}{R_2}\right) + Q,$$

$$P(t_0) = P_0, \tag{19}$$

where $P = P(t) = E[\hat{x}(t) - x(t)]^2$ is actual variance of the filtering error $\tilde{x}(t) = \hat{x}(t) - x(t)$.

Together with the optimal solution, the Kalman filter equations (18) and (19), we apply the suboptimal linear filter (5), (6), (10) and (11). We denote the estimates of variable $x(t)$ based on observations $\{y_1(\tau),\ t_0 \le \tau \le t\}$ and $\{y_2(\tau),\ t_0 \le \tau \le t\}$ by $\hat{x}_1$ and $\hat{x}_2$, respectively. Using the equations (5) and (6) for $k = 1, 2$, we obtain the equations for $\hat{x}_1$ and $\hat{x}_2$:

$$\dot{\hat{x}}_1 = a\hat{x}_1 + P_1 \frac{b_1}{R_1}(y_1 - b_1\hat{x}_1),$$

$$\hat{x}_1(t_0) = m_0, \tag{20}$$

$$\dot{P}_1 = 2aP_1 - P_1^2 \frac{b_1^2}{R_1} + Q,$$

$$P_1(t_0) = P_0, \tag{21}$$

and

$$\dot{\hat{x}}_2 = a\hat{x}_2 + P_2 \frac{b_2}{R_2}(y_2 - b_2\hat{x}_2),$$

$$\hat{x}_2(t_0) = m_0, \tag{22}$$

$$\dot{P}_2 = 2aP_2 - P_2^2 \frac{b_2^2}{R_2} + Q,$$

$$P_2(t_0) = P_0. \tag{23}$$

At $N = 2$, the Eqs. (11) have the form

$$c_1(P_{11} - P_{12}) + c_2(P_{21} - P_{22}) = 0,$$
$$c_1 + c_2 = 1 \quad (P_{12} = P_{21}) \tag{24}$$

Whence it follows that

$$c_1 = (P_{22} - P_{21})(P_{11} - P_{12} - P_{21} + P_{22})^{-1},$$
$$c_2 = 1 - c_1. \tag{25}$$

And after simple transformations, it is easy to show that the Eqs.(25) coincide with the Eqs. (9). The formulae (25) take the exact form

$$c_1 = (P_{22} - P_{12})/(P_{11} - 2P_{12} + P_{22}),$$
$$c_2 = (P_{11} - P_{12})/(P_{11} - 2P_{12} + P_{22}). \tag{26}$$

¿From (10) and (11) at $N = 2$, the suboptimal estimate $\hat{x}^*$ takes the form

$$\hat{x}^* = \frac{P_{22} - P_{12}}{P_{11} - 2P_{12} + P_{22}}\hat{x}_1 + \frac{P_{11} - P_{12}}{P_{11} - 2P_{12} + P_{22}}\hat{x}_2. \tag{27}$$

Now we calculate the actual variance of the filtering error $\Delta = \Delta(t) = \hat{x}^*(t) - x(t)$,

$$
\begin{aligned}
P^* &= E[\Delta^2] \\
&= \left(\frac{P_{22} - P_{12}}{P_{11} - 2P_{12} + P_{22}}\right)^2 P_{11} \\
&\quad + \frac{2(P_{22} - P_{12})(P_{11} - P_{12})}{(P_{11} - 2P_{12} + P_{22})^2}P_{12} \\
&\quad + \left(\frac{P_{11} - P_{12}}{P_{11} - 2P_{12} + P_{22}}\right)^2 P_{22}, \tag{28}
\end{aligned}
$$

where $P_{11} = P_1 = E[\Delta_1^2]$ and $P_{22} = P_2 = E[\Delta_2^2]$ are actual variances of the filtering errors $\Delta_1 = \hat{x}_1 - x$ and $\Delta_2 = \hat{x}_2 - x$, respectively, and $P_{12} = E[\Delta_1\Delta_2]$ is the cross-covariance function of the errors Δ_1 and Δ_2.

According to equation (16) at $i = 1$ and $j = 2$ the actual cross-covariance function P_{12} of the errors Δ_1 and Δ_2 is determined by the following equation:

$$\dot{P}_{12} = \left(2a - P_1\frac{b_1^2}{R_1} - P_2\frac{b_2^2}{R_2}\right)P_{12} + Q,$$

$$P_{12}(t_0) = P_0. \tag{29}$$

The formula (28) and the equations (21), (23) and (29) produce the actual accuracy of the suboptimal filter (20)-(23) and (27).

At $Q = 0.2$, $a = -1, b_1 = 1, b_2 = 2, R_1 = \frac{1}{3}$ and $R_2 = 0.5$ it is easy to show that the steady-state values of the variances $P_1(\infty)$, $P_2(\infty)$ and $P(\infty)$ are equal to 0.0883, 0.0765 and 0.0717, respectively. Putting $dP_{12}/dt = 0$ in Eq. (29), we find the steady-state value of correlation P_{12} as $P_{12}(\infty) = 0.0695$. Calculating the coefficients c_1 and c_2 by (26) we have $c_1 = 0.2713$, $c_2 = 0.7287$.

Next we calculate the steady-state value of variance $P^*(\infty)$ of dependent filtering error

$$\tilde{x} = x^* - x, \quad \hat{x}^* = c_1\hat{x}_1 + c_2\hat{x}_2. \tag{30}$$

Substituting the values $c_1 = 0.2713$, $c_2 = 0.7287$, $P_{11}(\infty) = P_1 = 0.0883$, $P_{22}(\infty) = P_2 = 0.0765$ and $P_{12}(\infty) = 0.0695$ into Eq. (15), we have

$$
\begin{aligned}
P^*(\infty) &= \sum_{i,j=1}^{2} c_i P_{ij} c_j \\
&= c_1^2 P_{11} + 2c_1 c_2 P_{12} + P_{22} c_2^2 \\
&= 0.0745. \tag{31}
\end{aligned}
$$

If we assume that estimation errors are uncorrelated. It is easy to show that $P^*(\infty) = 0.0755$.

Consequently, although the filtering algorithm (20)-(27) is not optimal, the filtering error variance P^* for this algorithm gives more accurate estimates than formula (14) does.

4.2 Normal Kalman filter with real data

The results presented here were produced by software implementation of the proposed track fusion algorithm using C++. In the experiment, target is modeled by the kinematic model (Robert A. Singer 1970)

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & \alpha \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} w(t), \qquad (32)$$

where $w(t)$ is white noise with variance $Q = 2$, α is the reciprocal of the maneuver time constant with value $\alpha = 1/20$ that means an evasive maneuver. Multiple sensor involves three sensors with its local processor(Kalman filter). The measurements at the three sensors are

$$y_k = [1\ 0\ 0]x + v_k, \quad k = 1, 2, 3. \qquad (33)$$

Here v_1, v_2 and v_3 are independent zero-mean gaussian white noises with variances $R_1 = 1.0, R_2 = 1.5$ and $R_3 = 2.0$, respectively. We used $x(0) = [1\ 1\ 0]^T$. The initial estimates were taken equal to $[1\ 1\ 0]^T$. The error covariance and cross error covariance are initialized with $P_{ij}(0) = 10I$, with I the identity and $i, j = 1,2$ and 3. The motivation for selecting a large initial error covariance is that this makes the estimates converge more quickly to the states. Error covariance is plotted in dB. Euler method was used with step size of 10 *msec*.

Let's define the error covariance P_i at $i = 1, 2$ and 3. Since P_i is symmetric, P_i is:

$$P_i = \begin{bmatrix} p1 & p2 & p3 \\ p2 & p4 & p5 \\ p3 & p5 & p6 \end{bmatrix}, \qquad (34)$$

with time-varying scalars.

Figure 2 presents the diagonal of the error covariances of P_1, P_2 and P_3. ¿From Figure 2, we know that showing the highest accurate estimates is $\hat{x}_1$, the next one is $\hat{x}_2$, and $\hat{x}_3$ shows the lowest accurate estimates. Now first, we combine two estimates,$\hat{x}_1$ and $\hat{x}_2$ and then combine three estimates, $\hat{x}_1$, $\hat{x}_2$ and $\hat{x}_3$ using Eqs. (11). Figure 3 shows the error covariance of track fusion of two estimates and track fusion of three estimates. From figure 3, we can see that track fusion of three estimates gives more accurate estimate than track fusion of two estimates does. It means that although $\hat{x}_3$ shows the lowest quality estimates, it might still contain useful information for tracking of the plant. Comparing figure 2 with figure 3 we see that the results of combining two sensor track is more accurate than each local sensor's estimates, and the results of combining three sensor track is more accurate than the results of combining two sensor track. The results show that the proposed filtering methods
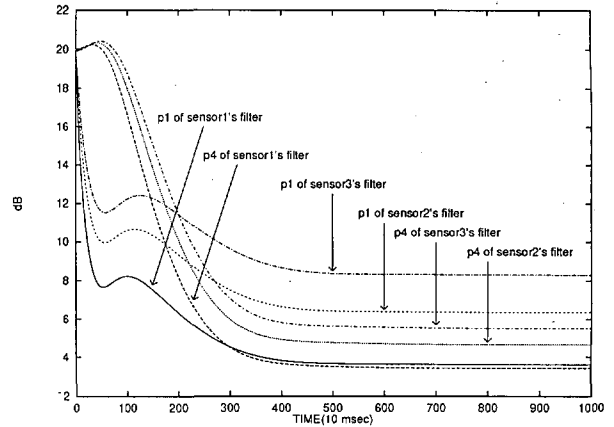

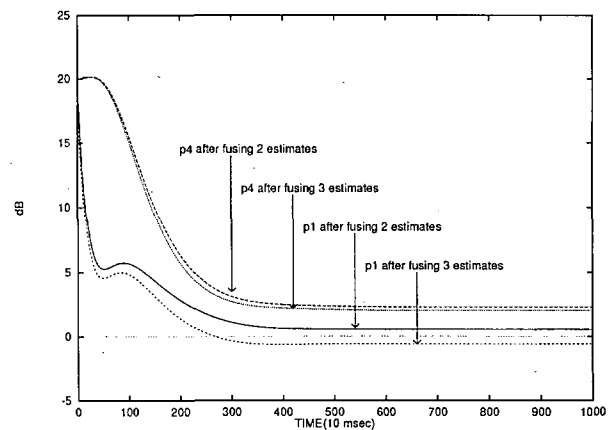Figure 2: Plots of the diagonals of the error covariance of P_1, P_2 and P_3

Figure 3: Comparing error covariance of track fusion of $\hat{x}_1$ and $\hat{x}_2$ with track fusion of $\hat{x}_1$, $\hat{x}_2$ and $\hat{x}_3$

can be used in multisensor environments to fuse local sensor's estimates to get more accurate estimates.

But peformance improvement when adding more sensor's information is limited. That means filtering error variance of proposed formula can not be less than the filtering error variance of formula (3). But the approach like formula (3) suffers primarily from the amout of observation data that must be transferred and central level failure(Bar-Shalom 1990).

5 Generalizing Fusion Formula to Discrete Filter

Recently signal processing is becoming digital. For this reason, it is important to generalize proposed fusion formula to discrete stochastic systems determined by difference equations.

5.1 Fusion formula in Discrete Filter

Proposed fusion formula can be generalized to discrete filter. Consider discrete stochastic system which state vector $x_k \in R^n$ is determined by linear difference equation:

$$x_k = F_{k-1}x_{k-1} + G_{k-1}w_{k-1}, \quad k = 1, 2, \ldots,$$

where k is discrete time, F_k and G_k are $n \times n$ and $n \times r$ matrices, respectively. The input noise $w_k \in R^r$ is gaussian random vector, $w_k \sim N(0, Q_k)$. We also assume that we have N different types of sensors determined by the following equations:

$$y_k^{(1)} = H_k^{(1)}x_k + v_k^{(1)}, \ldots, y_k^{(N)} = H_k^{(N)}x_k + v_k^{(N)}.$$

Here the observation error $v_k^{(i)}$ is gaussian random vector, $v_k^{(i)} \sim N(0, R_k^{(i)})$, $i = 1, \ldots, N$.

Then the new proposed suboptimal estimate $\hat{x}_{k|k}^*$ of the state vector x_k is constructed from the estimates $\hat{x}_{k|k}^{(1)}, \ldots, \hat{x}_{k|k}^{(N)}$ by the following fusion formula as we did in continuous filter:

$$\hat{x}_{k|k}^* = \sum_{i=1}^N c_k^{(i)}\hat{x}_{k|k}^{(i)}, \tag{35}$$

where

$$\sum_{i=1}^{N-1} c_k^{(i)}[P_{k|k}^{(ij)} - P_{k|k}^{(iN)}) + c_k^{(N)}[P_{k|k}^{(Nj)} - P_{k|k}^{NN}] = 0,$$

$$(j = 1, \ldots, N-1), \qquad \sum_{i=1}^N c_k^{(i)} = I$$

represents the linear algebraic Eqs. for matrices $c_k^{(1)}, \ldots, c_k^{(N)}$. $P_{k|k}^{(ij)}$ is cross-covariance matrix of the filtering errors $\tilde{x}_{k|k}^{(i)}$ and $\tilde{x}_{k|k}^{(j)}$ at $i \neq j$ and $P_{k|k}^{(ii)}$ is covariance matrix of the filtering error $\tilde{x}_{k|k}^{(i)}$, i.e.,

$$P_{k|k}^{(ij)} = E\left[\tilde{x}_{k|k}^{(i)}\tilde{x}_{k|k}^{(j)T}\right],$$

$$P_{k|k}^{(ii)} = P_{k|k}^{(i)} = E\left[\tilde{x}_{k|k}^{(i)}\tilde{x}_{k|k}^{(i)T}\right].$$

The equations for covariance matrices $P_{k|k}^{(i)}$ have the form (A. Gelb 1974):

$$P_{k|k}^{(i)} = E\left[I - K_k^{(i)}H_k^{(i)}\right]P_{k|k-1}^{(i)},$$

$$P_{1|0}^{(i)} = P_0.$$

And equations for the cross-covariance matrix $P_{k|k}^{(ij)}$ have the form (Y. Bar-Shalom 1981):

$$P_{k|k}^{(ij)} = [I - K_k^{(i)}H_k^{(i)}]F_{k-1}P_{k-1|k-1}^{(ij)}F_{k-1}^T$$
$$\cdot [I - K_k^{(j)}H_k^{(j)}]^T + [I - K_k^{(i)}H_k^{(i)}]$$
$$\cdot G_{k-1} \cdot Q_{k-1}G_{k-1}^T[I - K_k^{(j)}H_k^{(j)}]^T,$$

$$P_{0|0}^{(ij)} = P_0 \ \forall i \neq j, \ i, j = 1, \ldots, N,$$

where $K_k^{(i)}$ is Kalman gain matrix.

5.2 Accuracy of the proposed fusion formula

Now we derive the equation for actual covariance matrix

$$P_{k|k}^* = E\left(\tilde{x}_{k|k} \ \tilde{x}_{k|k}^T\right), \ \tilde{x}_{k|k} = \hat{x}_{k|k}^* - x_k, \tag{36}$$

where x_k is the state vector, and $\hat{x}_{k|k}^*$ is the proposed estimate(35).

Substituting (35) into (36) we have

$$P_{k|k}^* = \sum_{i,j=1}^N c_k^{(i)}P_{k|k}^{(ij)}c_k^{(j)T},$$

where the covariance matrices $P_{k|k}^{(ij)}$ are determined in previous section.

Thus we have completely define the new suboptimal linear filter for estimate $\hat{x}_{k|k}^*$ of the state vector x_k.

Remarks

1) The proposed estimate(35) is unbiased, i.e.,

$$E\hat{x}_{k|k}^* = E[\sum_{i=1}^N c_k^{(i)}E[\hat{x}_{k|k}^{(i)}]]$$

$$= [\sum_{i=1}^N c_k^{(i)}]Ex_k = Ex_k, \quad k = 0, 1, \ldots$$

2) We assume the block diagonal covariance matrix $R_k = diag[R_k^{(1)} \ldots R_k^{(N)}]$ of composite measurement noise $v_k = [v_k^{(1)T} \ldots v_k^{(N)T}]^T$ that means measurement noises are independent of each other. If R_k is not diagonal (for example, in the case of dependent noises), the *Cholesky decomposition method* (Mohinder S. Grewal and Angus P.Andrews 1993) can be applied to find linear transformation of the observation vector $y_k = [y_k^{(1)T} \ldots y_k^{(N)T}]^T$, whose measurement noise will be block diagonal.

6 Conclusion

In this paper, we introduced suboptimal filtering method for N-sensor track fusion with the common process noise and also generalize these methods to discrete stochastic system determined by difference equation. This method is based on observation decomposition method where the composite observation vector is decomposed to N different types of subsystem. Each sensor's estimate is fused by the minimum mean square error criterion. Theoretical and experimental results have been derived that show accuracy and effectiveness. The proposed suboptimal filtering method can be widely used in the different areas of applications: industrial, military, space, target tracking, inertial navigation and others (Ren C.Luo, Michael G.Kay 1989). This method reduces

the size of required computations by introducing full parallelism in the design procedures.

Appendix

Derivation of Equations (11). The estimation error has the form

$$\tilde{x} = \hat{x}^* - x = \sum_{k=1}^{N} c_k \hat{x}_k - x = \sum_{k=1}^{N} c_k (\hat{x}_k - x)$$

$$= \sum_{k=1}^{N} c_k \tilde{x}_k.$$

Let's compute the covariance P of error $\tilde{x}$. We have

$$P = E[\tilde{x}\tilde{x}^T] = \sum_{i,j=1}^{N} c_i E[\tilde{x}_i \tilde{x}_j^T] c_j^T$$

$$= \sum_{i,j=1}^{N} c_i P_{ij} c_j^T,$$

$$P_{ij} = E[\tilde{x}_i \tilde{x}_j^T]. \tag{37}$$

We seek the optimal matrices (coefficients) c_k $(k = 1, \ldots, N)$ minimizing the mean square error, i.e.,

$$\Phi = tr(P) \to min_{c_k}, \tag{38}$$

at the following conditions

$$c_1 + \cdots + c_N = I. \tag{39}$$

Next we transform the formula for P,

$$P = \sum_{i,j=1}^{N} c_i P_{ij} c_j^T$$

$$= \sum_{i,j=1}^{N-1} c_i P_{ij} c_j^T + \sum_{i=1}^{N-1} (c_i P_{iN} c_N^T + c_N P_{Ni} c_i^T)$$

$$+ c_N P_{NN} c_N^T. \tag{40}$$

Substituting the expression $c_N = I - c_1 \cdots - c_{N-1}$ into

(40) we have

$$P = \sum_{i,j=1}^{N-1} c_i P_{ij} c_j^T$$

$$+ \sum_{i=1}^{N-1} \{c_i P_{iN} (I - \sum_{j=1}^{N-1} c_j)^T\}$$

$$+ \sum_{i=1}^{N-1} \{(I - \sum_{j=1}^{N-1} c_j) P_{Ni} c_i^T\}$$

$$+ (I - \sum_{j=1}^{N-1} c_j) P_{NN} (I - \sum_{j=1}^{N-1} c_j)^T$$

$$= \sum_{i,j=1}^{N-1} c_i P_{ij} c_j^T + \sum_{i=1}^{N-1} (c_i P_{iN} + P_{Ni} c_i^T)$$

$$- \sum_{i,j=1}^{N-1} (c_i P_{iN} c_j^T + c_j P_{Ni} c_i^T) + P_{NN}$$

$$- (\sum_{i=1}^{N-1} c_i) P_{NN} - P_{NN} (\sum_{j=1}^{N-1} c_j)^T$$

$$+ \sum_{i,j=1}^{N-1} c_i P_{NN} c_j^T. \tag{41}$$

So, the optimization problem (38), (39) has the following form:

$$\Phi = tr(P) \to min_{c_k}, \quad k = 1, \ldots, N-1, \tag{42}$$

where the covariance matrix P is determined by the formula (41).

Next we'll use the following formulae:

$$P_{ij} = P_{ji}^T, \quad P_{ii} = P_{ii}^T \ for \ any \ i, j, \tag{43}$$

$$\frac{\partial}{\partial c_h} tr(c_h X c_h^T) = c_h X^T + c_h X,$$

$$\frac{\partial}{\partial c_h} tr(c_h X) = X^T,$$

$$\frac{\partial}{\partial c_h} tr(X c_h^T) = X. \tag{44}$$

Let's differentiate every summand of the funtional $\Phi =^{(41)} tr(P)$ with respect to c_h $(h = 1, \ldots, N-1)$ using (43) and (44) and then set the result to zero. We have

$$\frac{\partial}{\partial c_h} tr(\sum_{i,j=1}^{N-1} c_i P_{ij} c_j^T) = \sum_{i=1, i \neq h}^{N-1} \frac{\partial}{\partial c_h} tr(c_i P_{ih} c_h^T)$$

$$+ \sum_{j=1, j \neq h}^{N-1} \frac{\partial}{\partial c_h} tr(c_h P_{hj} c_j^T) + \frac{\partial}{\partial c_h} tr(c_h P_{hh} c_h^T)$$

$$= 2 \sum_{i=1, i \neq h}^{N-1} c_i P_{ih} + 2 c_h P_{hh}, \tag{45}$$

$$\frac{\partial}{\partial c_h} tr[\sum_{i=1}^{N-1} (c_i P_{iN} + P_{Ni} c_i^T)]$$

$$= \frac{\partial}{\partial c_h} tr(c_h P_{hN} + P_{Nh} c_h^T) = 2P_{Nh}, \qquad (46)$$

$$\frac{\partial}{\partial c_h} tr(\sum_{i,j=1}^{N-1} c_i P_{iN} c_j^T) = \sum_{i=1,i\neq h}^{N-1} \frac{\partial}{\partial c_h} tr(c_i P_{iN} c_h^T)$$

$$+ \sum_{j=1,j\neq h}^{N-1} \frac{\partial}{\partial c_h} tr(c_h P_{hN} c_j^T) + \frac{\partial}{\partial c_h} tr(c_h P_{hN} c_h^T)$$

$$= \sum_{i=1,i\neq h}^{N-1} c_i P_{iN} + \sum_{j=1,j\neq h}^{N-1} c_j P_{Nh}$$

$$+ c_h(P_{hN} + P_{Nh}), \qquad (47)$$

$$\frac{\partial}{\partial c_h} tr(\sum_{i,j=1}^{N-1} c_j P_{Ni} c_i^T) = \sum_{i=1,i\neq h}^{N-1} \frac{\partial}{\partial c_h} tr(c_h P_{Ni} c_i^T)$$

$$+ \sum_{j=1,j\neq h}^{N-1} \frac{\partial}{\partial c_h} tr(c_j P_{Nh} c_h^T) + \frac{\partial}{\partial c_h} tr(c_h P_{Nh} c_h^T)$$

$$= \sum_{i=1,i\neq h}^{N-1} c_i P_{iN} + \sum_{j=1,j\neq h}^{N-1} c_j P_{Nh}$$

$$+ c_h(P_{Nh} + P_{hN}), \qquad (48)$$

$$\frac{\partial}{\partial c_h} tr(\sum_{i=1}^{N-1} c_i P_{NN}) = \frac{\partial}{\partial c_h} tr(c_h P_{NN})$$

$$= P_{NN}, \qquad (49)$$

$$\frac{\partial}{\partial c_h} tr[(P_{NN}(\sum_{i=1}^{N-1} c_i)^T] = \frac{\partial}{\partial c_h} tr(P_{NN} c_h^T)$$

$$= P_{NN}, \qquad (50)$$

$$\frac{\partial}{\partial c_h} tr(\sum_{i,j=1}^{N-1} c_i P_{NN} c_j^T) = \sum_{i=1,i\neq h}^{N-1} \frac{\partial}{\partial c_h} tr(c_i P_{NN} c_h^T)$$

$$+ \sum_{j=1,j\neq h}^{N-1} \frac{\partial}{\partial c_h} tr(c_h P_{NN} c_j^T) + \frac{\partial}{\partial c_h} tr(c_h P_{NN} c_h^T)$$

$$= 2\sum_{i=1,i\neq h}^{N-1} c_i P_{NN} + 2c_h P_{NN}. \qquad (51)$$

So, by the formulae (45) - (51) the Eq. $\frac{\partial}{\partial c_h} tr(P) = 0$

has the form

$$\sum_{i=1,i\neq h}^{N-1} c_i P_{ih} + c_h P_{hh} + P_{Nh} - \sum_{i=1,i\neq h}^{N-1} c_i P_{iN}$$

$$- \sum_{j=1,j\neq h}^{N-1} c_j P_{Nh} - c_h(P_{hN} + P_{Nh}) - P_{NN}$$

$$+ (\sum_{i=1,i\neq h}^{N-1} c_i) P_{NN} + c_h P_{NN} = 0,$$

$$h = 1,\dots,N-1,$$

or

$$\sum_{i=1,i\neq h}^{N-1} c_i(P_{ih} - P_{iN}) + (P_{Nh} - P_{NN})$$

$$+ c_h(P_{hh} - P_{Nh} - P_{hN} + P_{NN})$$

$$- (\sum_{j=1,j\neq h}^{N-1} c_j)(P_{Nh} - P_{NN}) = 0,$$

$$(h = 1,\dots,N-1). \qquad (52)$$

Substituting the expression

$$\sum_{j=1,j\neq h}^{N-1} c_j = {}^{(39)} I - c_N - c_h$$

into (52) we have

$$\sum_{i=1,i\neq h}^{N-1} c_i(P_{ih} - P_{iN}) + (P_{Nh} - P_{NN})$$

$$- (I - c_N - c_h)(P_{Nh} - P_{NN})$$

$$+ c_h(P_{hh} - P_{Nh} - P_{hN} + P_{NN}) = 0,$$

$$h = 1,\dots,N-1. \qquad (53)$$

And after simple transformations the Eqs. (53) take the form

$$\sum_{i=1}^{N-1} c_i(P_{ih} - P_{iN}) + c_N(P_{Nh} - P_{NN}) = 0,$$

$$h = 1,\dots,N-1. \qquad (54)$$

So, the set of Eqs. (54) together with Eq. $c_1 + c_2 + \cdots + c_N = I$ represents the linear algebraic Eqs. for matrices $c_1,\dots,c_N$.

References

[1] A. Gelb. (1974) Applied Optimal Estimation, MIT Press, Cambridge, Massachusetts, 1974 p 110.

[2] Mohinder S.Grewal and Angus P.Andrews. (1993) Kalman Filtering, Prentice-Hall,Inc, 1993 p 214-225.

[3] Ren C. Luo, Michael G. Kay. (1989) Multisensor Integration and Fusion in Intelligent Systems. *IEEE Trans. Systems, Man, and Cybernetics*, Vol. 19, No.5 1989, p. 901-931.

[4] Robert A. Singer. (1970) Estimating Optimal Tracking Filter Performance for Manned Maneuvering Targets. *IEEE Transactions on Aerospace and Electronic System*, VOL. AES, NO 4 July. 1970, p. 473-483.

[5] Yonggun Cho., et al.(1996) Suboptimal Continuous Fitering Based on The Decomposition of Observation Vector. *Computers Math. Applic.*, Vol. 32 No. 4, pp. 23-31, 1996.

[6] Y. Bar-Shalom. (1981) On the Track-to-Track Correlation Problem. *IEEE Transactions on Automatic Control*, Vol. AC-26. NO.2,APRIL 1981, p. 571-572.

[7] Y. Bar-Shalom and Leon Campo. (1986) The Effect of the Common Process Noise on the Two-Sensor Fused-Track Covariance. *IEEE Transactions on Aerospace and Electronic Systems*, AES-22(Nov.1986), P. 803-805.

[8] Y. Bar-Shalom (1990) Multitarget-Multisensor Tracking, Artech House, INC, 1990 pp 191-198.

The ROL Deductive Object-Oriented Database System

Mengchi Liu
Department of Computer Science
University of Regina
Regina, Saskatchewan Canada S4S 0A2
Email: mliu@cs.uregina.ca

ROL is a deductive object-oriented database system developed at the University of Regina. It supports important object-oriented features such as object identity, complex objects, classes, class hierarchies, multiple inheritance with overriding and blocking, and schema definition. It also supports structured values such as functor objects and sets, providing powerful mechanisms for representing both partial and complete information about sets. It provides a uniform declarative language for defining, manipulating and querying databases. This paper describes the structure and operation of ROL.

1 Introduction

Deductive object-oriented databases are intended to combine the best of the deductive and object-oriented approaches, such as recursion, declarative querying, and a firm logical foundation from the deductive approach, and object identity, complex objects, typing, classes, class hierarchies, multiple inheritance, and schema definition from the object-oriented approach. In the past decade, a number of deductive object-oriented database languages have been proposed, such as O-logic [16], revised O-logic [11], IQL [2], F-logic [10], LOGRES [6], LLO [15], Complex [8], Noodle [17], CORAL++[21], DLT [3], Gulog [7], and Rock & Roll [4].

However, most of these proposals stay at the theoretical level. Some of them, such as F-logic, are technically too complicated and it is far from clear how they could be fully implemented efficiently and taken as the basis of practical database systems. On the other hand, some implemented systems such as CORAL++, Rock & Roll, lack declarative querying and formal logical semantics but provide imperative languages.

In the last several years, we have designed and implemented a novel deductive object-oriented database system called ROL. The ROL system provides a uniform declarative language also called ROL [12, 13] for defining, manipulating and querying databases. It supports important object-oriented features such as object identity, complex objects, typing, classes, class hierarchies, multiple inheritance with overriding and blocking, and schema definition. In addition, it also supports structured values such as functor objects and

sets, treating them as first class citizens, and providing powerful mechanisms for representing both partial and complete information about sets. As a result, the ROL language is not only an object-oriented language, but also an extension of the traditional deductive database language Datalog (with negation) and subsumes it as a special case, which makes it quite different from other deductive object-oriented database languages. Indeed, ROL can be used for pure object-oriented deductive databases or for pure value-oriented deductive databases. Most importantly, ROL allows both value-oriented and object-oriented features to be used together to take advantage of both approaches.

In ROL, values, object identifiers, functor objects and sets are treated uniformly as objects so that functional dependencies can be represented directly and more generally than in functional data models [9, 19]. ROL builds-in other important integrity constraints such as domain, key, referential and cardinality in a simple and uniform framework.

ROL provides powerful set representation mechanisms that combine LDL [18, 5] and F-logic [10] set treatments. Information about sets can be represented partially or completely. It has a uniform notion of schema for objects represented both extensionally and intensionally. Most importantly, it has a well-defined logical semantics that cleanly accounts for most of its object-oriented and value-oriented features [13].

ROL supports schema queries, and facts and rule queries in addition to traditional database queries. It also supports declarative updates of schema, facts and rules.

The ROL system has been implemented as a single-

user database system. It runs on SUN workstations under Solaris, SGI workstation under IRIX, and DEC workstations under Ultrix, and PC compatible machines under Linux. Since its first public release in January of 1996, it has been downloaded by hundreds of users world-wide and used for various data and knowledge-based applications. It has also been used to teach advanced database courses on deductive databases, object-oriented databases, and deductive object-oriented databases at a number of universities around the world. We are still continuously enhancing the system, adding functionality and increasing performance.

In this paper, we describe the structure and operation of ROL. This paper is organized as follows. Section 2 introduces the ROL databases. Section 3 describes the architecture of the current ROL implementation. Sections 4, 5, and 6 discuss ROL data definitions, data queries, and data manipulations respectively. Section 7 outlines some future directions for our project.

2   ROL Databases

In this section, we first discuss classes and objects in ROL. Then we introduce ROL databases.

2.1   Classes

In ROL, classes and objects are distinguished and classes are not objects. Classes denote collections of objects that share the same set of attributes. An object in the collection denoted by a class is an instance of the class. There are four kinds of classes in ROL:

(1) value classes such as *integer, real, string,* $integer(0..5), string(\{'M', 'F'\})$;

(2) object identifier classes such as *person, student, course, supplier, part*;

(3) functor classes such as *supplies(supplier, part), student_course(student, course), family(person, person)*;

(4) set classes such as $\{person\}, \{course\}, \{part\}$.

Classes are related to each other through aggregation and generalization [20] using class declarations of the following form:

$$c[a_1 \Rightarrow c_1, ..., a_n \Rightarrow c_n] \; isa \; d_1, ..., d_m$$

where $c, c_1, ..., c_n, d_1, ..., d_m$ are classes, and $a_1, ..., a_n$ are attributes with $n \geq 0, m \geq 0$. It defines that the class c has directly defined attributes $a_1, ..., a_n$ on the classes $c_1, ..., c_n$ and is an immediate subclass of $d_1, ..., d_m$. The class c inherits all attribute definitions from its superclasses $d_1, ..., d_m$ but can also override

or block them and introduce additional attributes local to itself. Besides, instances of the class c are also instances of its superclasses $d_1, ..., d_m$.

The following are examples of class declarations:

$person[age \Rightarrow integer(0..125),$
$\qquad gender \Rightarrow string(\{'M', 'F'\}),$
$\qquad father \Rightarrow person,$
$\qquad mother \Rightarrow person]$

$student[age \Rightarrow integer(15..30),$
$\qquad takes \Rightarrow \{course\}] \; isa \; person$

$employee[age \Rightarrow integer(18..65),$
$\qquad salary \Rightarrow integer] \; isa \; person$

$workstudent \; isa \; student, person$

The class *person* has attributes *age, gender, father* and *mother*. The classes *student* and *employee* are immediate subclasses of the class *person*. They both inherit all attribute definitions other than *age* from the class *person*, but override (refine) the attribute definition for *age* and introduce an additional attribute definition local to themselves. The class *workstudent* is an immediate subclass of the classes *student* and *employee* and multiply inherits attribute definitions from both superclasses by taking their conjunction.

The following examples demonstrate how to block attribute inheritance from superclasses in ROL:

$orphan[father \Rightarrow none,$
$\qquad mother \Rightarrow none] \; isa \; person$
$french \; isa \; person$
$french\_orphan \; isa \; french, orphan$

The class *orphan* is defined as a subclass of *person*. The use of the built-in class *none* blocks the inheritance of the attribute definitions for *father* and *mother* from *person* to *orphan* and its subclass *french_orphan*.

In ROL, not only object identifier classes but also value classes, functor classes, and set classes can have attributes. Consider the following examples:

$integer[factorial \Rightarrow integer]$
$supplies(supplier, part)[$
$\qquad quantity \Rightarrow integer(0..500)]$
$\{person\}[count \Rightarrow integer]$

The value class *integer* has an attribute *factorial*, the functor class *supplies(supplier, part)* has an attribute *quantity*, and the set class has an attribute *count*.

2.2   Objects

Corresponding to classes, four kinds of objects are distinguished:

(1) values such as integers 5, 8, reals 3.14, 1.0, and strings *'Bob', 'Ann'*;

(2) object identifiers such as pam, bob, s_1, p_1;

(3) functor objects such as $supplies(s_1, p_1)$, $hobbies(bob, \{tv\})$;

(4) sets such as $\{\}, \{pat\}, \{hobbies(bob, \{tv\})\}$.

Objects have attributes through which they are related to each other. ROL is a typed language, objects and their attribute values must be well-typed with respect to the attribute definitions of their direct classes.

Facts and rules in ROL provide the extensional and intensional information about objects and their attribute values, respectively. A fact is represented using one of the following forms in ROL:

$$v[a_1 \rightarrow o_1, ..., a_n \rightarrow o_n]$$
$$o : c[a_1 \rightarrow o_1, ..., a_n \rightarrow o_n]$$
$$f(o'_1, ..., o'_m)[a_1 \rightarrow o_1, ..., a_n \rightarrow o_n]$$
$$\{o'_1, ..., o'_m\}[a_1 \rightarrow o_1, ..., a_n \rightarrow o_n]$$

where v is a value, o is an object identifier, $o'_1, ..., o'_m$ are objects, c is an object identifier class, f is the name of a functor class, $a_1, ..., a_n$ are attributes, and $o_1, ..., o_n$ are objects or partial sets to be introduced shortly, and $m > 0, n \geq 0$. The first one specifies that the value v has values $o_1, ..., o_n$ for its attributes $a_1, ..., a_n$. The second one specifies that the object identifier o is a direct instance of the class c and has values $o_1, ..., o_n$ for its attributes $a_1, ..., a_n$. The third one specifies that the functor object $f(o'_1, ..., o'_n)$ has a primary class f with values $o_1, ..., o_n$ for its attributes $a_1, ..., a_n$. The last one specifies that the set $\{o'_1, ..., o'_n\}$ has values $o_1, ..., o_n$ for its attributes $a_1, ..., a_n$.

The following are examples of ROL facts:

$1[factorial \rightarrow 1]$
$bob : person[age \rightarrow 25,$
$\qquad\qquad father \rightarrow tom,$
$\qquad\qquad mother \rightarrow pam]$
$supplies(s_1, p_1)[quantity \rightarrow 200]$
$\{jim, pat\}[count \rightarrow 2]$

The first fact says that the factorial of 1 is 1. The second one says that the object identifier *bob* is a direct instance of the class *person* and has age 25, father *tom* and mother *pam*. The next one says that shipment identified by the functor object $supplies(s_1, p_1)$ has quantity 200. The last one says that count of the set $\{jim, pat\}$ is 2.

ROL is a rule-based system. Information about objects and their attribute values can be represented using rules intensionally in ROL. In order to allow the value of a set-valued attribute to be inferred using rules in more than one step, the notion of partial set is introduced. Like a set which is also called complete set, a partial set is a collection of objects represented with different notation. For example, $\langle jim, pat \rangle$ is a partial set. Unlike a complete set, a partial set is not an object and therefore cannot have attributes. It is used to denote part of an object (a complete set). For example, $\langle jim, pat \rangle$ denotes part of a set that contains *jim*

and *pat* which could be $\{jim, pat\}$ or $\{jim, pat, sam\}$ depending on the context. Partial sets can only be used as partial attribute values.

The following facts show how to represent information about the value of a set-valued attribute *ancestors* of *bob* partially:

$$bob[ancestors \rightarrow \langle tom, pam \rangle]$$
$$bob[ancestors \rightarrow \langle sam \rangle]$$

The first fact says that *tom* and *pam* are ancestors of *bob* and the second says that *sam* is also an ancestor of *bob*. Neither of them directly says what *bob*'s ancestors exactly are. However, if nothing else about *bob*'s children can be obtained from the database, then the above two facts are equivalent to the following fact:

$$bob[ancestors \rightarrow \{tom, pam, sam\}]$$

A rule is an expression of the form:

$$A :- L_1, ..., L_n$$

where the head A is an object expression and the body $L_1, ..., L_n$ is a sequence of object expressions, negated object expressions, usual arithmetic and set-theoretic comparison expressions. An object expression is a fact with variables occurring in some places of objects, attributes and classes.

The following rules show how to derive partial attribute values using rules.

$$X[parents \rightarrow \langle Y \rangle] :- X[father \rightarrow Y]$$
$$X[parents \rightarrow \langle Y \rangle] :- X[mother \rightarrow Y]$$
$$X[ancestors \rightarrow \langle Y \rangle] :- X[parents \rightarrow \langle Y \rangle]$$
$$X[ancestors \rightarrow \langle Y \rangle] :- X[ancestors \rightarrow \langle Z \rangle],$$
$$Z[parents \rightarrow \langle Y \rangle]$$

The first two rules say that if Y is the father or the mother of X, then $\langle Y \rangle$ is a partial value of the parents of X. The next two rules specify how to derive partial values for the attribute *ancestors* of X recursively. The ROL system automatically combines intermediate partial attribute values of objects.

The following rules show how to use negation and arithmetic and set-theoretic comparison expressions to derive attribute values.

$$X[trueAncs \rightarrow \langle Y \rangle] :- X[ancestors \rightarrow \langle Y \rangle],$$
$$\neg X[parents \rightarrow \langle Y \rangle]$$
$$X[factorial \rightarrow Y] :- X_1[factorial \rightarrow Y_1],$$
$$X = X_1 + 1,$$
$$Y = Y_1 \times X$$
$$S[count \rightarrow X] :- S = S_1 \cup S_2, S_1 \cap S_2 = \{\},$$
$$S_1[count \rightarrow Y],$$
$$S_2[count \rightarrow Z],$$
$$X = Y + Z$$

The first rule says that if Y is one of the ancestors of X and Y is not one of the parents of X, then Y is one of the true ancestors of X. The second rule says

that if the factorial of X_1 is Y_1, X is $X_1 + 1$, then the factorial of X is $Y_1 \times X$. The last rule says that if the set S can be partitioned into two subsets S_1 and S_2 such that the count of S_1 is Y and the count of S_2 is Z, then the count of S is $Y + Z$.

2.3 Databases

A ROL database consists of three parts: a schema, a set of facts and a set of rules. The schema is a set of class declarations. The facts and rules are the extensional and intensional information about objects, their direct classes and/or their attribute values respectively.

Example 1 The following is an example of ROL database.

Schema
$$person[father \Rightarrow person,$$
$$mother \Rightarrow person,$$
$$parents \Rightarrow \{person\},$$
$$ancestors \Rightarrow \{person\},$$
$$trueAncs \Rightarrow \{person\}]$$

Facts
$$tom : person$$
$$pam : person$$
$$bob : person[father \rightarrow tom, mother \rightarrow pam]$$
$$liz : person[mother \rightarrow pam]$$
$$jim : person[father \rightarrow bob]$$
$$pat : person$$
$$sam : person[father \rightarrow jim, mother \rightarrow pat]$$
Rules
$$X[parents \rightarrow \langle Y \rangle] :- X[father \rightarrow Y]$$
$$X[parents \rightarrow \langle Y \rangle] :- X[mother \rightarrow Y]$$
$$X[ancestors \rightarrow \langle Y \rangle] :- X[parents \rightarrow \langle Y \rangle]$$
$$X[ancestors \rightarrow \langle Y \rangle] :- X[ancestors \rightarrow \langle Z \rangle],$$
$$Z[parents \rightarrow \langle Y \rangle]$$
$$X[trueAncs \rightarrow \langle Y \rangle] :- X[ancestors \rightarrow \langle Y \rangle],$$
$$\neg X[parents \rightarrow \langle Y \rangle]$$

The set of rules in a ROL database is required to be stratified. Stratification is automatically determined by using a dependency graph which is a marked graph constructed as follows:

1. The set of nodes consists of all classes and attributes that appear in or can substitute for class and attribute variables in rules in the database.

2. There is an edge from x to y if there is a rule in which x is in the head and y is in the body. The edge is marked if y is in an negated object expression or y is in an object expression $P[y \rightarrow Q]$ such that Q is not of the form $\langle Q_1, ..., Q_n \rangle$ and there exists a rule such that $P'[y \rightarrow \langle Q'_1, ..., Q'_m \rangle]$ is in its head.

The dependency graph for the rules in the above database is shown as follows.
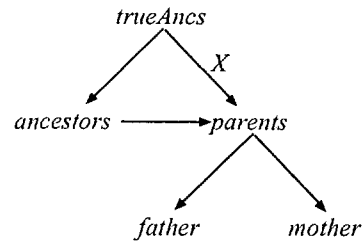
The set of rules in a database is stratified if the dependency graph has no cycle with a marked edge. For example, the set of rules in Example 1 is stratified as the only marked edge is not in a cycle.

The dependency graph is not only used by the ROL system to determine stratification but also to decide how to evaluate the rules upon queries and updates.

Example 2 The following is another example of a ROL database.

Schema
$$part[name \Rightarrow string]$$
$$basepart[cost \Rightarrow integer,$$
$$mass \Rightarrow integer] \ isa \ part$$
$$compart[subparts \Rightarrow \{part\},$$
$$assemcost \Rightarrow integer,$$
$$massincr \Rightarrow integer] \ isa \ part$$
$$\{part\}[totalcost \Rightarrow integer,$$
$$totalmass \Rightarrow integer]$$
Facts
$$p_1 : basepart[cost \rightarrow 200, mass \rightarrow 30]$$
$$p_2 : basepart[cost \rightarrow 100, mass \rightarrow 20]$$
$$p_3 : basepart[cost \rightarrow 150, mass \rightarrow 20]$$
$$p_4 : basepart[cost \rightarrow 300, mass \rightarrow 40]$$
$$p_5 : compart[subparts \rightarrow \{p_2, p_3\}]$$
$$p_6 : compart[subparts \rightarrow \{p_1, p_5\}]$$
$$p_7 : compart[subparts \rightarrow \{p_4, p_6\}]$$
Rules
$$\{X\}[totalcost \rightarrow Y, totalmass \rightarrow Z] :-$$
$$X[cost \rightarrow Y, mass \rightarrow Z]$$
$$\{X\}[totalcost \rightarrow Y, totalmass \rightarrow Z] :-$$
$$X[assemcost \rightarrow Y, massincr \rightarrow Z]$$
$$S[totalcost \rightarrow Y, totalmass \rightarrow Z] :-$$
$$S = S_1 \cup S_2, S_1 \cap S_2 = \{\},$$
$$S_1[totalcost \rightarrow Y_1, totalmass \rightarrow Z_1],$$
$$S_2[totalcost \rightarrow Y_2, totalmass \rightarrow Z_2],$$
$$Y = Y_1 + Y_2, Z = Z_1 + Z_2$$
$$X[assemcost \rightarrow Y, massincr \rightarrow Z] :-$$
$$X[subparts \rightarrow S],$$
$$S[totalcost \rightarrow Y, totalmass \rightarrow Z]$$

This is a manufacturing company's parts database. There are two kinds of parts: base parts and composite parts. The class *part* has a single-valued attribute

name. Both classes *basepart* and *compart* are subclasses of *part* and inherit the *name* attribute from *part*. The class *basepart* has single-valued attributes *cost* and *mass*. *compart* has a set-valued attribute *subparts* whose values are sets of parts and single-valued attributes *assemcost* and *massincr*. The set class {*part*} has single-valued attributes *totalcost* and *totalmass*.

The facts in the database are about the base parts and their cost and mass as well as composite parts and the way they are made from other parts.

The rules are used to compute the *totalcost* and *totalmass* of sets of parts and the *assemcost* and *massincr* of composite parts. The first rule states that if the part X has *cost* Y and *mass* Z, then the singleton set {X} has *totalcost* Y and *totalmass* Z. The second rule is similar. The third rule specifies that in order to compute *totalcost* and *totalmass* of a set of parts, first partition the set into two subsets and use the rule recursively and then combine the results. The last rule says that the *assemcost* and *massincr* of a composite part can be obtained by using the *totalcost* and *totalmass* of the *subparts* set.

The dependency graph for the rules in the above database is shown as follows:
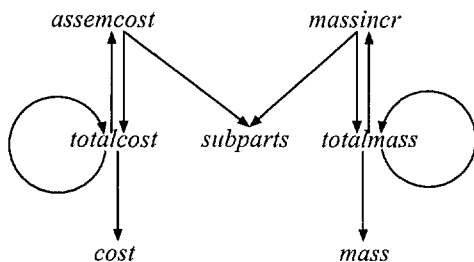
As there is no marked edge in the graph, the set of rules is stratified.

In [1], a set of four tasks is proposed to evaluate the expressiveness of database programming languages using the above manufacturing company's parts database. The four tasks are:

(1) Describe the database

(2) Print the name and cost of all base parts that cost more than $100

(3) Compute the total cost of a given composite part

(4) Record the decomposition of a new part in the database, i.e., how a new composite part is manufactured from its subparts

It is clear that the first and third tasks can be accomplished successfully in ROL. We will show how to perform the second and last tasks in Sections 5 and 6 respectively.

3 Architecture Overview

The architecture of the ROL system is shown in Figure 1.
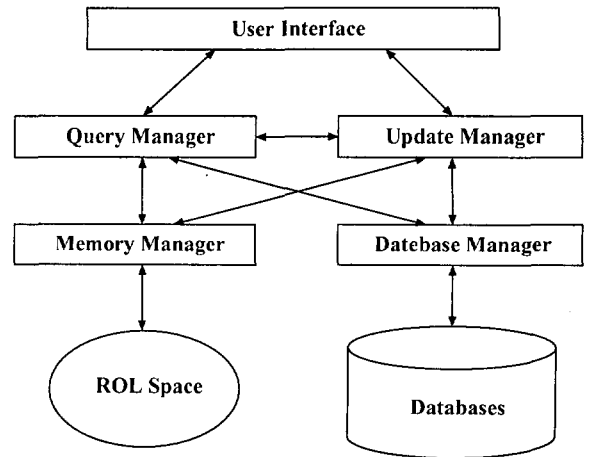
Figure 1: Architecture of the ROL system

The system is organized into three layers. The first layer is the user interface. Two kinds of user interfaces are provided:

1. textual user interface that interacts with the user by reading in user commands and processing their answers

2. graphic user interface that allows the user to select operations from the menus

Both kinds of interfaces provide different kinds of environment for the user to access the database. They perform syntactical analysis of the user query or update request, filter out output processing command, rewrite the request into standard form, send the possibly modified request to the query or update manager, and process the answers to the request received from the lower layer based on the output processing command they keep.

The second layer processes the query or update request from the user interface. In ROL, an update request may imply a query. Thus the update manager may send the query to query manager before performing the update.

The memory manager manages the ROL space in memory that is used to store meta-information about facts, rules and class definitions, part of facts, rules, and class definitions that need to be used to process queries, and intensional information derived with rules.

The database manager provides rapid access to facts, rules, class definitions and meta information about them on the disk.

The ROL system supports persistent facts, rules and class definitions. The user can query facts, rules, and

class definitions in their original forms or intensional information that can be derived from them. Computations are driven by user query requests and are done in the ROL space.

In order to speed up query answering and avoid redundant computation, the ROL system keeps part of intensional information derived with rules in the ROL space. However, intensional information is not made persistent as extensional information. The reason we make such a decision is that if we make intensional information persistent, update propagation may be ineffective on disk. By only allowing updates to extensional information, keeping part of intensional information in the ROL space gives the best trade-off in terms of performance.

Therefore, the ROL system uses rules only if the intensional data that can be derived with rules is not in the ROL space.

If the memory is used up, the memory manager removes the least recently used data systematically from the ROL space and modify the corresponding meta information about them.

4   Data Definitions

The ROL system provides the following commands for data definitions.

createdb	*destroydb*
open	*close*
insert	*delete*

The *createdb* command is used to create a new ROL database. The *destroydb* command is used to destroy an existing ROL database. The ROL system allows multiple databases to exist on disk. However, only one database can be accessed at a time. The database to be accessed must be opened first by the user with the *open* command. To switch to another database, the user has to close the opened (or current) database with the *close* command. The *insert* and *delete* commands are used to add, delete, or modify the schema of the database. Data definition requests are handled by the update manager.

The following example shows how to create the schema of a ROL database called *part* shown in Example 2.

(1) *createdb part*
(2) *open part*
(3) *insert part[name ⇒ string]*
(4) *insert basepart[cost ⇒ integer] isa part*
(5) *insert compart[subpart ⇒ {part},*
    assemcost ⇒ integer] isa part
(6) *insert {part}[cost ⇒ integer]*

The following example shows how to modify the schema of the *parts* database.

(1) *delete part[name ⇒ _]*
(2) *delete basepart[cost ⇒ _],*
    insert basepart[cost ⇒ integer(0..2000]
(3) *delete {part}*
(4) *delete part*

The first operation deletes the directly defined attribute *name* of *part* as well as the values of the attribute *name* of the instances of *part*. The second one modifies the attribute definition of *cost*. It will fail if some instances of *basepart* violate the new attribute definition. The third one deletes the class *{part}* and all its instances from the database as well as any attribute definitions that involve it such as *subparts* in Example 2. The last one deletes the classes *class* and *{part}* as well as their instances. Furthermore, all functor classes and attribute definitions that involve them and the corresponding attribute values will be deleted.

5   Data Queries

Before the user can query a database, the database must be opened first. By opening the database, the ROL system loads into the ROL space meta information about facts, rules and class definitions.

The ROL system provides the *query* command to query ROL databases.

Three kinds of queries are supported in the ROL system: schema queries, object queries, fact and rule queries. The ROL system always gives all answers at once (i.e., set-at-a-time).

Schema Queries Schema queries are used to retrieve information about classes, their attribute definitions, and subclass relationships. The user defines the immediate subclass relationships with *isa* in the schema and general subclass relationships *isa∗* between classes can then be derived and queried. The following are several schema queries:

(1) *query X*
(2) *query X isa part*
(3) *query compart[A ⇒ X]*
(4) *query X[A ⇒ integer]*

The first query asks for every class defined in the schema. The second one asks for every immediate subclass of *part*. The third one asks for every attribute definition of the class *compart*. Only the attributes that are directly defined or inherited but not overridden or blocked will be given in the result. The last one asks for every class and its attribute that is defined on *integer*.

Object Queries The user can query the information about objects represented extensionally by facts

or intensionally by rules in the database. The result to such a query is always based on the final evaluated and combined information. The following are several object query examples:

(1) *query X : part*
(2) *query X[name →, Y, cost → Z], Z > 100*
(3) *query p_7[assemcost → X]*
(4) *query {p_4, p_6}[totalcost → X]*

The first query asks for every instance of the class *part*. The second one asks for the name and cost of all base parts that costs more than 100. This is actually the second task shown in Section 2. The third one asks for the assembly cost of the composite part p_7. The last one asks for the totalcost of the set {p_4, p_6}.

Fact and Rule Queries The object queries allow us to obtain evaluated and combined information about objects which is represented intensionally and/or extensionally. Sometimes, we need to know how the information about objects is represented in the database. That is, we need to query facts and rules in their original form. Such queries are supported in ROL. Consider the following examples:

(1) *query p_1[cost]*
(2) *query [cost]*
(3) *query [totalcost]*
(4) *query compart[ ]*

The first query asks for part of the fact that gives the value of the attribute *cost* of p_1. The second one asks for part of the facts that give the value for the attribute *cost* of every base part. The third one asks for all the rules that are used to derive the values of the attribute *totalcost*. The last query asks for all the rules that are used to derive the attribute values of instances of the class *compart*.

Query Output Handling In ROL, query results are displayed in a special way. For example, the results for *query X[cost → Y]* will be displayed on the screen as

$X = p1, Y = 20$
$X = p2, Y = 10$
...

For many applications, this kind of display of results may not be the way the user likes. For this reason, ROL provides the *write* command to handle query results. The user can use it to organize the query results as in Pascal and C++. For example, the following query

query X[cost → Y],

write(stdout, X,' [cost →', Y,']')

will display the results on the screen as follows:

p1[cost → 20]
p2[cost → 10]
...

Furthermore, the query results can also be stored in a file called *test* as follows:

query X[cost → Y],
write(test, X,' [cost →', Y,']')

The *write* command can also perform the projection on query results. That is, if the variables appear in the query but not in the write command, then the values for the variables will be discarded.

Query Processing Based on the user query, the ROL system uses the meta information in the ROL space to decide which facts, rules, and class definitions should be loaded into the ROL space and which facts and rules can be removed from the ROL space if memory space is needed. It also keeps track of what is in the ROL space. It then uses one of the following four kinds of evaluation strategies to process the user query:

(1) matching
(2) semi-naive bottom-up evaluation
(3) top-down evaluation without general unification
(4) matching/bottom-up and top-down alternations

Each strategy is particularly efficient for some cases, but may not be applicable to or perform relatively poorly on others. The ROL system automatically decides which mechanism to use without user's intervention, based on the nature of the query and its knowledge about the data in the database. The details about query processing can be found in [14].

6   Data Manipulations

The *insert* and *delete* commands used for data definitions are also used for data manipulations in ROL. That is, they are used to manipulate facts and rules in the database. Furthermore, they can be used together with the *query* command to perform complex updates. Data manipulations are handled by the update manager as well.

When a new fact is to be inserted, the update manager will first check if it is well-typed by calling query manager for class definition. Only if the fact is well-typed, the update manager will then insert it into the database.

The following examples show how to insert facts and rules in Example 2 into the database:

(1) *insert* p_1 : *basepart*[*cost* → 20]
(2) *insert* p_2 : *basepart*[*cost* → 10]
(3) *insert* p_3 : *basepart*[*cost* → 15]
(4) *insert* $\{X\}$[*totalcost* → Y] :- X[*cost* → Y]

The following example shows how to accomplish the last task shown in Section 2:

$$insert\ p_8 : compart[subparts \rightarrow \{p_5, p_6, p_7\}]$$

The *insert* and *delete* commands can also be used to delete part of existing facts and update existing facts. Consider the following examples:

(1) *delete* X[*name* → _]
(2) *delete* p_7
(3) *query* X[*assemcost* → Y], Y > 50, *delete* X
(4) *delete* X[*cost* → Y], Y > 20, Y' = Y + 10,
    insert X[*cost* → Y']

The first operation deletes the name of every part. The second one deletes the object identifier p_7 and all of its attribute values from the database completely. It is also deleted from all attribute values where it appears so that the referential integrity constraints are maintained. The third one deletes each composite part that has assembly cost higher than 50 from the database completely. The last one increases the cost of the base part that costs more than 20 by 10.

Now let us see how the update manager process data manipulations. Consider the following operation again:

$$delete\ X[cost \rightarrow Y], Y > 20, Y' = Y + 10,$$
$$insert\ X[cost \rightarrow Y']$$

The update manager first generate the following four parts:

$$query\ X[cost \rightarrow Y], Y > 20$$
$$Y' = Y + 10$$
$$delete\ X[cost \rightarrow Y]$$
$$insert\ X[cost \rightarrow Y']$$

It then executes the first part and finds all the bindings for variables X and Y. Then it computes the bindings for Y' based on the bindings of Y. With the bindings for all variables in insertion and deletion expressions, it can perform the specified operations. In order to achieve atomicity of update operations, ROL checks if the insertions could cause any problems such as type errors for the object to be inserted before it actually performs any operations, If so, it just stops and reports associated errors. Otherwise, ROL performs deletions both on disk and in the ROL space and propagates the deletions to intensional information in the ROL space. Finally, it performs insertions on disk and in the ROL space. Therefore, ROL supports declarative updates

of facts and rules. That is, the order of operations in an update request is immaterial so that the update request has a declarative meaning.

The ROL system also supports the deletion of rules. Rule queries can be used in the *delete* command. Consider the following examples:

(1) *delete* $\{X\}$[*totalcost* → Y] :- X[*cost* → Y]
(2) *delete* $R2$
(3) *delete* [*totalcost*]
(4) *delete* *compart*[ ]

The first operation deletes the specified rule. In the ROL system, each rule is assigned a rule number automatically which can be found by rule queries. Instead of writing the complete rule out, the user can simply use the rule number in the command. Thus, if $R2$ is associated with a rule in the database, then the second operation deletes that rule. The third one deletes all rules that can be found with the rule query [*totalcost*], that is, all rules that are used to derive the values of the attribute *totalcost*. The last one deletes all rules that can be found with the rule query *compart*[ ], that is, all rules that are used to derive the attribute values of the instances of the class *compart*.

The insertion of rules may cause rules in the database no longer stratified. When inserting a rule, ROL will ensure stratification restriction not violated and generate evaluation strategies for different cases of the rules using data dictionary information about stratification and evaluation strategies for the existing rules. When a rule is deleted from the database as well as the ROL space, data dictionary is modified accordingly and all intensional information relevant to the rule is also deleted from the ROL space.

7   Conclusions

This paper has described the structure and operation of ROL, a disk-based deductive object-oriented database system under continuous development at the University of Regina.

Four versions have been released so far. The latest version is available over the Internet from the World Wide Web home page at `http://www.cs.uregina.ca/~mliu/ROL`.

As ROL is only structurally object-oriented rather than behaviorally object-oriented, we would like to add behavioral aspects to it. We are also working on various query optimization mechanisms to enhance its performance. The client-server based multiple user system is also underway.

Acknowledgments

mentation of the ROL system: Weidong Yu, Min Guo, Riqiang Shan, Minqing Xing and Jun Shu. This work has been partly supported by the Natural Sciences and Engineering Research Council of Canada.

References

[1] S. Abiteboul and C. Beeri. On the power of languages for the manipulation of complex objects. In *Proceedings of the International Workshop on Theory and Applications of Nested Relations and Complex Objects in Databases*, Darmstadt, Germany, 1987. Springer-Verlag LNCS 361.

[2] S. Abiteboul and P. C. Kanellakis. Object identity as a query language primitive. In *Proc. ACM SIGMOD Intl. Conf. on Management of Data*, pages 159–173, Portland, Oregon, 1989.

[3] R. Bal and H. Balsters. A deductive and typed object-oriented language. In S. Ceri, K. Tanaka, and S. Tsur, editors, *Deductive and Object-Oriented Databases*, pages 340–359, Phoenix, Arizona, USA, 1993. Springer-Verlag LNCS 760.

[4] M. L. Barja, A. A. A. Fernandes, N. W. Paton, M. H. Williams, A. Dinn, and A. I. Abdelmoty. Design and implementation of ROCK & ROLL: a deductive object-oriented database system. *Information Systems*, 20(3):185–211, 1995.

[5] C. Beeri, S. Naqvi, O. Shmueli, and S. Tsur. Set construction in a logic database language. *J. Logic Programming*, 10(3,4):181–232, 1991.

[6] F. Cacace, S. Ceri, S. Crepi-Reghizzi, L. Tanca, and R. Zicari. Integrating object-oriented data modelling with a rule-based programming paradigm. In *Proc. ACM SIGMOD Intl. Conf. on Management of Data*, pages 225–236, 1990.

[7] G. Dobbie and R. Topor. On the Declarative and Procedural Semantics of Deductive Object-Oriented Systems. *Journal of Intelligent Information System*, 4(2):193–219, 1995.

[8] Sergio Greco, Nicola Leone, and Pasquale Rullo. COMPLEX: An object-oriented logic programming system. *IEEE Transactions on Knowledge and Data Engineering*, 4(4):344–359, 1992.

[9] B. C. Housel, V. Waddle, and S. B. Yao. The functional dependency model for logical database design. In *Proc. Intl. Conf. on Very Large Data Bases*, pages 194–208, Rio De Janeiro, Brazil, 1979. Morgan Kaufmann Publishers, Inc.

[10] M. Kifer, G. Lausen, and J. Wu. Logical foundations of object-oriented and frame-based languages. *Journal of ACM*, 42(4):741–843, 1995.

[11] M. Kifer and J. Wu. A logic for programming with complex objects. *J. Computer and System Sciences*, 47(1):77–120, 1993.

[12] M. Liu. An overview of the ROL language. In *Proc. Intl. Symp. on Cooperative Database Systems for Advanced Applications (CODAS '96)*, Kyoto, Japan, December 5-7 1996. World Scientific Publishing Co.

[13] M. Liu. ROL: A deductive object base language. *Information Systems*, 21(5):431 – 457, 1996.

[14] M. Liu and W. Yu. Query processing in the ROL system. In *Proc. Intl. Database Engineering and Applications Symp (IDEAS '97)*, Montreal, Canada, 1997. IEEE-CS Press.

[15] Y. Lou and M. Ozsoyoglu. LLO: A deductive language with methods and method inheritance. In *Proc. ACM SIGMOD Intl. Conf. on Management of Data*, pages 198–207, Denver, Colorado, 1991.

[16] D. Maier. A logic for objects. Technical Report CS/E-86-012, Oregon Graduate Center, Beaverton, Oregon, 1986.

[17] I. S. Mumick and K. A. Ross. Noodle: A language for declarative querying in an object-oriented database. In S. Ceri, K. Tanaka, and S. Tsur, editors, *Deductive and Object-Oriented Databases*, pages 360–378, Phoenix, Arizona, USA, 1993. Springer-Verlag LNCS 760.

[18] Shamim Naqvi and Shalom Tsur. *A Logical Language for Data and Knowledge Bases*. Computer Science Press, 1989.

[19] D. W. Shipman. The functional extending the database relational model to capture more meaning. *ACM Trans. on Database Systems*, 4(4):297–434, 1979.

[20] J. M. Smith and D. C. P. Smith. Database abstraction: Aggregation and generalization. *ACM Trans. on Database Systems*, 2(2):105–133, 1977.

[21] D. Srivastava, R. Ramakrishnan, D. Srivastava, and S. Sudarshan. Coral++: Adding object-orientation to a logic database language. In *Proc. Intl. Conf. on Very Large Data Bases*, pages 158–170, Dublin, Ireland, 1993. Morgan Kaufmann Publishers, Inc.

Conscious Representations, Intentionality, Judgements, (Self)Awareness and Qualia

Mitja Peruš
National Institute of Chemistry, Lab. for Molecular Modeling and NMR
Hajdrihova 19 (POB 3430), SI-1001 Ljubljana, Slovenia
fax: ++386-61-1259-244, E-mail: `mitja.perus@uni-lj.si`

The problem of consciousness is discussed by considering and commenting on theories of mental representations, judgements, feelings and a subject's self-aware qualitative experience. These characteristics of consciousness are correlated with materialistic background processes, except qualia which have, as it is argued, no obvious connection with naturalistic explanatory framework.

1 Introduction

Recently, the problem of *qualia*, i.e. the nature and origin of qualitative subjective experience, "how it is to be like" (the first-person perspective), and the problem of phenomenal consciousness (as opposed to its informational aspect) have been identified as the "hard" problem (Chalmers, 1995; Tucson II, 1996). The problems of system-dynamical background of consciousness and information processing (the naturalistic, third-person perspective) have, on the other hand, been defined as the "easy" problems, because they are not as puzzling as the phenomenal qualia (Banks, 1996; Hubbard, 1996).

Searle (1993), as a philosopher with non-reductive views (as opposed to eliminativist materialists like the Churchlands), defines consciousness as a subjective qualitative process of awareness, sentience or feeling. We can assume that consciousness has a multiple nature which is at the same time synthesized into a unity: multi-modal perceptions and representations are always unified into a single undividable experience (binding problem). Self-referential dynamics is essential for consciousness, but it can be understood in the framework of *multi-level neural, quantum or/and subcellular network processes* which obey *analogous collective dynamics* (Peruš, 1997a,e). On the other hand, qualitative phenomenal experience can not be described using conventional naturalistic, functionalist and information-theoretical models. Therefore some new, multi-disciplinary and non-reductionist approaches are necessary.

From the *naturalistic* point of view (Burnod, 1990; Pribram, 1991; Nagel, 1993; Baars, 1997; Newmann, 1997), multi-level coherence of various fractal-like complex biosystems is necessary for cognition and consciousness. Similar collective information processes in neural and quantum networks, which remind one of holography, support this view (Peruš, 1997b,c). The most important neuro-quantum analogy, beside the interference-based memorization, is that the reconstruction of a neuronal pattern (the recall of a pattern from memory) is analogous to the so-called "wave-function collapse". In the neural case, from a superposition of neuronal patterns one pattern alone is made clear in the system of neurons (representing the object of consciousness), all the others remain implicitly stored in the system of synaptic connections (in memory) only. In a quantum system, the wave-function "collapses" from a superposition of eigen-wave-functions to a state which can be described by a single eigen-wave-function, all the others are latent, enfolded in the implicate order (Peruš, 1995c, 1996). These processes provide a processual background for consciousness, and can explain bidirectional *consciousness–memory transitions* as well as *unconsciousness–consciousness transitions*. There is an interesting parallelism between making a pattern or thought conscious and transforming quantum implicit or potential states into explicit or actually existent states. This is also related to the "arrow of time" (Hameroff et al., 1996).

2 Transitional mental representations

Intentional consciousness needs mental representations in order to represent the objects we are conscious of. In the usual neural network models (Amit, 1989; Peretto,

1992), they are approximated by neuronal patterns-qua-attractors (Peruš, 1995a,b). These models may (or even should) be used in a generalized manner – as networks of *formal* "neurons". These formal "units" can be implemented in various media, e.g. neural, quantum, virtual, etc. (Peruš, 1995c).

Existence of mental representations was in the centre-point of discussions between cognitivists (representationalists) and ecologists (e.g., J.J. Gibson). Ecologists emphasize the importance of the environment and its continuous mutual interactions with the organism which unites the organism and its environment into an indivisible dynamical whole.

But the question remains whether specific external patterns are projected into specific internal *representations* (cognitivists' position), or does the brain only make *specific transformations* of specific input patterns to a specific output, i.e. a specific response of the organism to the stimuli from the environment (ecologists' position). In the first case representations (pictorial, propositional, linguistic) are more fixed and stable. They have well-defined semantic kernels which are relatively independent from their context. In the second case "representations" would be purely dynamic and transitional, with strong dependence on the context.

In connectionist models, there are two main levels – cognitive (symbolic) and sub-cognitive (sub-symbolic) level. It must be emphasized that they coexist and that there are also many quasi-levels in-between (McClelland et al., 1986). The connectionist level constitutes an underlying system-processual medium for high-order mental representations – in the same way that quantum physics is a necessary processual background for classical physics. The invariance of cognitive representation-patterns is only an "envelope" for very complex internal dynamics on the sub-cognitive or sub-symbolic levels.

The neural medium is able to "absorb" each external pattern in order to "get into its shape", as Aristotle would say. Each external pattern is not constantly represented in the brain, but only when the influence from environment forces it to reconstruct the corresponding representation. All un-interesting or un-important patterns are only abstractly coded in the system of synaptic connections and wait there for reconstruction or "unfolding" when necessary. The brain makes a combination of environmental influences and of the use of representational codes in the memory, i.e. pattern-correlations encoded in the system of synaptic connections. It makes a superposition of external patterns and already stored internal codes (which represent the organism's expectations). The environmental influence selectively extracts those features from the memory which are the most similar to the actual state in the environment. [1] So, mental representations get into correlative coherence with external patterns, or, with other words, mental processes get into parallel synchronous dynamics with environmental processes. Computer simulations (e.g., Peruš, 1995a) also show that neural network dynamics depends very much on the structure (e.q., correlation structure) of the input data.

The conclusion would be that there *are* some *strongly environment-dependent representations* in the mind, but these are not static at all – they are *very dynamic, flexible* and *adaptive*, carrying only the *filtered (abstracted) main characteristics* of the patterns.

The structure of internal representations (and their semantic relation-network) is an isomorphic virtual image of the structure of environment (Peruš, 1995b), including its individual patterns, their spatial and temporal correlations, and groups of environmental patterns. Brentano says (Brentano, 1973, pp. 9) that our spatial and temporal world exhibits the same relations as those exhibited by the object of our perceptions of space and time.

To summarize, associative neural network models and their computer simulations support the view that epistemic intermediaries exist, but are strongly environment-dependent, dynamic, and only *transitional*. Phantom limbs offer evidence that virtual representations exist – injured people, after the amputation of their limb, continue to feel that they still have it. For this feeling, transitional (temporary) virtual intermediaries are responsible. Such intermediaries are not states, they are processes. The rate of transitionality varies inverse-proportionally to the rate of stability, invariance and importance of external patterns and their internal representational counterparts, to their frequency of occurence, and to the amount of attention paid to them.

3 Informational background of phenomena

Specific mappings of specific objects into specific mental representations (i.e., specific patterns of neural activity acting as *attractors*) constitute the *intentional and representational basis of phenomenal experience* (Sajama et al., 1987), but cannot explain the *qualitative* nature of phenomena. Therefore I will postpone the discussion of qualitative component of phenomena, and now consider merely their informational (so-called "access"-consciousness) component (Davies & Humphreys, 1993) (Marcel & Bisiach, 1988).

From a purely informational view-point, a phenomenon is an object perceived "through" the state of the neural system (Nelkin, 1996). In all phenomena, objects and their representations are always "bound

---

[1]Detailed descriptions of these processes, based on author's computer simulations, are given in (Peruš, 1995a,b, 1997a).

together". Both, objects (as far as they are phenomena to us) and their representations, have no meaning or no existence one without the other. Furthermore, phenomena represent correlation or coherence, or even effective unification of objects with their mental representations. Brentano says that objects of sensations are merely phenomena, and that color, sound, warmth, taste etc. do not really exist outside our sensations, even though they may point to objects which do so exist (Brentano, 1973, pp. 9). He also says (ibid., pp. 69) that color is not seeing and sound is not hearing. We could say that color is a characteristics of an object (only when this object acts as a phenomenon to us!) only as much as it was gained through the process of seeing. Namely, we might have uncoupled object and neural system initially.

Objects and "their" characteristics (color etc.) can be phenomena only when they are perceived (seen etc.). So, seeing always effectively unifies everything that is denoted by the notions of object, phenomenon and its color into a virtual whole. Neural system must be coupled with the object through this process of seeing. This unification is only virtual and effective one – as it would establish a sort of higher-order meta-gestalt which compounds the mental (virtual, emergent) and the physical (system-processual) into one – into a "virtual unity". So, we must speak not about a man and an object separately, but about a man-seeing-an-object (Sajama et al., 1987).

Phenomena do not have properties like shape and size, but they do posses analogues to those properties. Phenomenal states are not coloured, yet they correlate to colours: their variations (including varying scales, intensity, etc.) are "read out" in such a way that we are led to judge a conclusion that a property of external objects varies in a similar way (Nelkin, 1996). In such a way, phenomena act as image-like qualitative representations. An epiphenomenalist would say that they have no role in perception itself, but are co-effects of the processes that result in percepts. They somewhat indirectly accompany perception and alter in parallel with the external objective situation. On the other hand, non-reductionists would attribute a causal role to phenomena. I would say that the latter are right, except in automatic (e.g., reflex) behaviour which is triggered before the irreducible conscious control of the subject's I is switched on and may, somewhat later, freely alter the actions.

4 Intentionality, judgements and emotional attitudes

According to Brentano, we never only think, but we think about something. This is, in philosophy, called *intentionality* (sometimes a more direct word – "aboutness" – is used). A pattern of neural activity is "car-

rier" of a specific content (correlated with an external pattern – object).

Due to Brentano (1973, pp. 278), representations, judgements and emotional attitudes are three basic, but interdependent, classes of mental reference. Brentano (1973, pp. 265) writes: "The inner consciousness, which accompanies every mental phenomenon, includes a presentation, a cognition and a feeling, all directed towards that phenomenon." Later (Brentano, 1973, pp. 276): "Every mental activity is the object of a presentation included within it and of a judgement included within it, it is also the object of an emotional reference included within it". "Nothing can be judged, desired, hoped or feared, unless one has a presentation of that thing."

Using the neural network theory, we can describe how patterns get associatively connected, because they (like their constituent neurons) are connected to each other and represent the context and content of each other. A representation is not only connected with other representations, but can also be symbolically represented (coded) by the firing of a cardinal neuron or a cardinal ensemble of neurons, or virtually by so-called order parameters (Haken, 1991). The firings of cardinal neurons symbolizes the occurence of their corresponding representations.

Judgements are intentional, even volitional, psychical events. They have neural correlates which may in our model be realized as "flipping" of cardinal neurons (or changing order parameters), which codes the mean-field situation ("general atmosphere", average) of the neural system and its global transitions. Judgements are affirmations or denials. In our model, the essential neural correlates of judgements may be represented by an excitatory or inhibitory action of a cardinal neuron towards corresponding pattern which is the object of judgement. The strength of activity of a cardinal neuron symbolizes the degree of conviction with which judgement is made.

Here we should add that in biological neural networks special "veto"-cells exist, which are protagonists of processes underlying judgements. But the role of "veto"-cells is limited to the context of system-dynamics only, i.e. the system triggers them. So, nothing *volitional* can be traced on the neural level. Our free will exists merely on the irreducible subjective level.

In a larger sense, emotional activity is love or hate which is correlated with higher-order pattern-agreement (mutual supporting) or disagreement. This causes convenience (pleasure etc.) or inconvenience (suffering etc.). The system-dynamic mechanisms underlying judgements (how they arise and what mental effects they have) and emotions can be modeled by neural networks, but the involvement of consciousness in the sense of free will, self-awareness and qualia cannot be satisfactorily understood this way.

5 Intentional consciousness and self-awareness

Intentionality means that every mental process always has a reference to a content or is directed upon an object (phenomenon). This is particularly characteristic of consciousness (except in transcendental mystical states which are un-intentional). According to Brentano (1973), intentionality represents a typically psychical phenomenon which cannot be reduced to physical phenomena, so it is an example of essential difference between the psychical and the physical. [1]

Dennett adds that action is intentional only if the actor is aware of action spontaneously ("automatically", without observation of the action). His example (Dennett, 1969, pp. 165): If somebody is tapping in the rythm of "Rule, Britannia" and is not aware of this (other people recognize this), then such tapping-in-a-rythm is not intentional.

Before we start to discuss the problem of consciousness and self-awareness (consciousness of consciousness) from the points of view of Brentano and of neural network theory, we have to emphasize tight connections and inter-dependence between representations and re-representations (Oakley, 1990). Brentano (1973, pp. 127) claims that there is a special connection between the object of inner representation and the representation of the representation, and so on. Let us quote Brentano (1973, pp. 127):
"The presentation of the sound and the presentation of the presentation of the sound form a single mental phenomenon; it is only by considering it in its relation to two different objects, one of which is a physical phenomenon and the other a mental phenomenon, that we divide it conceptually into two presentations. In the same mental phenomenon, in which the sound is presented to our minds, we simultaneously apprehend the mental phenomenon itself. What is more, we apprehend it in accordance with its dual nature insofar as it has the sound within it, and insofar as it has itself as content at the same time."

Brentano's introspective observations remain valid also if we understand the term "(re)presentations" as "neuronal patterns-qua-attractors". Namely, neuronal patterns are "reflecting each other" (heteroassociation) or are "reflecting themselves (into themselves)" (autoassociation) (Peruš, 1995a). *Neurons and patterns consisting of neurons represent context and con-*

tent to each other, and patterns represent context and content to themselves within themselves, because the neurons, which constitute them, are constantly interacting.

The recursive "self-intentionality" is the basis of the process of self-awareness. Memory consists of superpositions of correlation-patterns in the system of synaptic connections. The object of consciousness is represented in the pattern of the system of neurons which is involved in a global associative connection or interplay with many other stored patterns. Awareness and self-awareness might correspond to Brentano's discussion of reflective re-representations (although Brentano did not explicitly mention self-awareness). Brentano (1973, pp. 128) continues:
"If an inner presentation were ever to become inner observation, this observation would be directed upon itself.
One observation is supposed to be capable of being directed upon another observation, but not upon itself. The truth is that something which is only the secondary object of an act can undoubtedly be an object of consciousness in this act, but cannot be an object of observation in it. Observation requires that one turns his attention to an object as a primary object. (...) Thus we see that no simultaneous observation of one's own act of observation or of any other of one's own mental acts is possible at all. We can observe the sounds we hear, but we cannot observe our hearing of the sounds. On the other hand, when we recall a previous act of hearing, we turn toward it as a primary object, and thus we sometimes turn toward it as observers. In this case, our act of remembering is the mental phenomenon which can be apprehended only secondarily."

Thus, Brentano offers a representative of a "copy-theory" of self-awareness. A man can be aware of a copy or recalled image of a just-passed-away mental event, but not of this mental event directly.

We must note that there is one exception to the exclusion of simultaneous representations and re-representations: experiences of mystical unity – insofar they are *un-intentional* (Raković & Koruga, 1996; Peruš, 1997d). They correspond to *coherent* symmetrical dynamics of the neural substrate on biological level and global-attractor-formations on higher virtual levels (pattern-superpositions or simultaneously coexisting representations merge into an uniform whole). Quantum correlates (e.g., Bose-Einstein condensates) of such processes are very probably also relevant.

Brentano admits the complementarity of the first-order consciousness and the accompanying second-order consciousness (1973, pp. 129):
"The consciousness of the presentation of the sound clearly occurs together with the consciousness of this consciousness, for the consciousness which accompa-

---

[1] The second difference is, according to Brentano (1973, pp. 85), that all physical phenomena have extension and spatial location, but mental phenomena (thinking, willing etc.) appear without extension and spatial location. On the other hand, quantum physics and parallel-distributed complex systems show that this division can be melted away. It is namely only a result of being-inside-an-attractor (extension, localization) or being-beyond-local-attractors ("flowing freely" across the set of possible system's states).

nies the presentation of the sound is a consciousness not so much of this presentation as of the whole mental act in which the sound is presented, and in which the consciousness itself exists concomitantly. Apart from the fact that it presents the physical phenomenon of sound, the mental act of hearing becomes at the same time its own object and content, taken as a whole."

We shall conclude with Brentano (1973, pp. 134) that, if we see a color and have a representation of our act of seing, the color which we see is also present in the representation of this act. This color is the content of the representation of the act of seeing, but it also belongs to the content of seeing. It is well known that there are good candidates (they incorporate self-interactive dynamics) of correlates of these self-reflective mental processes on the levels of neural, sub-cellular and/or neural networks arising from iterative fractal-like dynamics (Peruš, 1997b).

6 Consciousness entailing self-consciousness

The relation of consciousness and self-consciousness was discussed in detail recently by Rosenthal and Gennaro. Gennaro (1995), following the higher-order-thought theory of (self)consciousness by Rosenthal (e.g., in Davies & Humphreys, 1993, pp. 197-224), shows that consciousness entails self-consciousness. [1]

Gennaro (1995) argues that a mental state S becomes conscious if it is accompanied by a meta-psychological thought M that one is in that mental state S. Examples of meta-psychological states are thoughts, beliefs, desires, wishes, hopes, fears. But not all such second-order states can render a first-order mental state conscious - it must be a meta-psychological *thought* directed at the first-order state, Gennaro says. "Self-consciousness does involve an explicit (albeit unconscious) thought and accompanies all conscious experience." (Gennaro, 1995, pp. 18)

Introspection is having a conscious thought about own mental state. Having conscious thought does not entail introspective awareness. In conscious states we are often not consciously thinking about our own thoughts, but we are nevertheless thinking about or having (unconscious) "thought awareness of" them. So, introspection is a special form of self-consciousness which includes conscious, not ordinary, higher-order thoughts. A meta-psychological state M is conscious

when a higher meta-psychological thought MM is directed on it. In introspection, a mental state S is accompanied by M, and, furthermore, there is a MM directed at M. Deliberate introspection is the most sophisticated kind of a S-M-MM loop.

One can have a particular phenomenal state without its typical quality, like in the following case. One can have muscle twinges while sleeping, and that causes changing position (Gennaro, 1995, p. 8). Thus, unconscious phenomenal states are possible and they may share (because they are *potentially* conscious) certain underlying neural processes with conscious phenomenal states.

Gennaro identifies the following conscious states: *conscious phenomenal states* (conscious bodily sensations, e.g. pains, and conscious world-directed perceptual states), *conscious world-directed non-perceptual intentional states* (e.g., desires, thoughts), and *self-consciousness* (non-reflective self-consciousness, i.e. unconscious meta-psychological thought awareness; momentary focused introspection and deliberate introspection).

Gennaro concludes that mental states (e.g. beliefs) *per se* do not require self-consciousness. However, episodic memories, self-directed thoughts and self-modification of behavior are needed for consciousness and they entail some form of self-consciousness. However, the nature of meta-thought remains an open question, at least partly because of the unsolved qualia problem.

The *I (ego)* as a proposition-like self-representation can be treated as an attractor or gestalt of the highest order (as far as its phenomenal character could be neglected – strictly it could not). Deep meditators can transcend their *I*s (egos) as soon as the corresponding global attractor is erased (see also Deikman, 1996). However, *I* is an irreducible trigger of volitional actions. These are connected with subjective motivation. And here again, qualia come into play, because of having an important role in motivation.

7 Qualia make consciousness an aspect beyond physicalism

To conclude, we will consider a characteristic attempt by a non-reductionist philosopher (N. Nelkin) to discuss (self)consciousness with phenomenal qualia. Phenomenal consciousness is a mental state with subjective feelings, i.e. something it is like to be in that state. It is essential for sensations like colourful visual and soundful auditory experiences, kinaesthetic feelings, pains. On the other hand, there is nothing it is like to be a conscious thinking itself, or a conscious feeling itself (Nelkin, 1996). So, if one is aware of his phenomenal quale, this is because one has a second-order, noninferential, proposition-like aware-

---

[1] The English notion "consciousness" cannot be well translated into many other languages, because it is used in a relatively broad sense. For example, German "das Bewußtsein", or "zavest", "svijest", "svest" in South-Slavic languages, which are usually translated as "consciousness", have implicit meaning of "self-consciousness" (much more than in English). This situation is in agreement with Gennaro's welcome, and in English context not trivial, idea that consciousness entails self-consciousness.

ness that one is in that first-order or phenomenal mental state. The awareness of qualitative states (making these qualia conscious) is an apperceptive second-order thought or even judgement, and as such this second-order state is more important for personality and self-identity than qualia (first-order phenomenal states) themselves. Nelkin (1996) distinguishes three types of conscious states: *first-order proposition-like representational state* (C1), *high-level, neurally-based, image-like representational state with phenomenality* (CS), and *second-order, direct, noninferential accessing and proposition-like representation of some C1 and of some CS* (C2).

CS is already a kind of phenomenal awareness, but C2 is a distinct apperceptive awareness (a real self-awareness).

Consciousness, in spite of its relative primitivity, un-analysability, un-effability and phenomenal unity (Kihlstrom, 1993), has several aspects, even unconscious implicit ingredients. Many beliefs we are apperceptively conscious of do not seem tied to phenomena. For example, one can be conscious that one believes tomorrow is Friday, but no set of phenomena is required for that consciousness (Nelkin, 1996). Secondly, phenomenal experience may alter because of physiological change. Patients with implanted new lenses complain that their colour phenomena are different.

It seems that non-phenomenal (i.e. "access-conscious", "purely" informational) states are those which, when directed toward C1 or CS, make the subject aware of them (C1 or CS). Nevertheless, in spite of not being universal and omnipresent, phenomenal qualitative states remain the central mystery (Hubbard, 1996). What really is the hardest problem is not awareness or self-directed awareness itself, but the qualitative nature of (self-)awareness. That is to say, with a theory of higher-order propositional-like thoughts we can somewhat trace the (cybernetic, recursive, iterative) essence of self-awareness (see also Železnikar, 1990), but we have no way to explain their phenomenal character. In spite of the fact that Nelkin tried to consider qualia a little bit more than Gennaro and Rosenthal, he did not succeed - as did not any other consciousness researcher (Flanagan, 1992; Hendriks-Jansen, 1997).

In this paper, selected theories of representations (following a connectionist line), intentionality (following Brentano's tradition) and consciousness together with self-reflective consciousness, i.e. self-awareness (starting with Brentano, then following Rosenthal and Gennaro) were discussed. At the end, the qualia enigma was stressed as the central and still unsolved problem.

Brain processes are merely a "centre of weight" of intentional consciousness. A non-local (sub)quantum coherent background, as transcendental mystical experiences suggest, is the origin of non-intentional consciousness or collective unconscious. The qualia problem, I believe, shows that consciousness is irreducible to neural, quantum and any other network processes, including their emergent, higher-level virtual, or informational processes, respectively. So, consciousness seems to be a *primary and irreducible aspect*, as primary as material processes. *A sort of "pre-consciousness" emerges in a primitive and rudimentary way as soon as complex material systems arise* (e.g., in sub-quantum "vacuum"), because *every complex system is accompanied by virtual attractor structures*. In a human-like form, however, consciousness (especially intentional consciousness) has evolved in the last millenia.

Pre-consciousness (rudimentary consciousness) is not prior to matter and matter is not prior to pre-consciousness. They both emerge together as soon as the fundamental sub-quantum symmetry is broken. On the other hand, our *phenomenal world* (i.e., objects *we* perceive) is a co-product of our intentional consciousness and of environment. We never know what the real world is, i.e. as it exists (probably it does) on its own, without us perceiving it. We only know how the world looks after it has been processed by our brains and our consciousness. In that sense, (intentional) consciousness is not prior to any thing-in-itself (Kant's "Ding an sich"), but it is prior to our phenomenal perception of a thing or object. There are no phenomena without (intentional) consciousness, but there may perhaps be things-in-themselves without it. However, pre-consciousness as well as matter are both essentially connected with overall sub-quantum background processes as their common "origin".

Because the origin of qualia is still unknown, and because of dependence on intersubjective definition of what consciousness is, the nature of consciousness remains a matter of hypotheses.

8   Acknowledgements

References

[1] Amit, D. (1989): Modeling Brain Functions (The world of attractor neural nets). Cambridge Univ. Press, Cambridge.

[2] Baars, B.J. (1997): In the Theater of Consciousness. Oxford Univ. Press, New York.

[3] Banks, W.P. (1996): How much work can a quale do? Consciousness & Cognition 5, 368-380.

[4] Brentano, F. (1973): Psychology from an Empirical Standpoint. Routledge & Kegan Paul, London. (German original: Psychologie vom empirischen Standpunkt, 1874.)

[5] Burnod, Y. (1990): An Adaptive Neural Network: the Cerebral Cortex. Prentice Hall, London.

[6] Chalmers, D.J. (1995): The puzzle of conscious experience. Scientific American (December), 62-68.

[7] Davies, M. & G.W. Humphreys (Eds.) (1993): Consciousness. Blackwell, Oxford.

[8] Deikman, A.J. (1996): "I" = awareness. J. Consciousness Studies 3, 350-356.

[9] Dennett, D.C. (1969): Consciousness and Content. Routledge & Kegan Paul, London.

[10] Flanagan, O. (1992): Consciousness Reconsidered. MIT Press, Cambridge (MA).

[11] Gennaro, R.J. (1995): Consciousness and Self-consciousness (A Defence of the Higher-Order Thought Theory of Consciousness). John Benjamins, Amsterdam / Philadelphia.

[12] Haken, H. (1991): Synergetic Computers and Cognition. Springer, Berlin etc.

[13] Hameroff, S.R.; A.W. Kaszniak & A.C. Scott (Eds.) (1996): Toward a Science of Consciousness - Tucson I. MIT Press, Cambridge (MA).

[14] Hendriks-Jansen, H. (1997): Information and the dynamics of phenomenal consciousness. Informatica 21, 389-404.

[15] Hubbard, T.L. (1996): The importance of a consideration of qualia to imagery and cognition. Consciousness & Cognition 5, 327-358.

[16] Kihlstrom, J.F. (1993): The continuum of consciousness. Cognition & Consciousness 2, 334-.

[17] Marcel, A.J & E. Bisiach (Eds.) (1988): Consciousness in Contemporary Science. Clarendon Press, Oxford.

[18] McClelland, J.L.; D.E. Rumelhart & PDP research group (1986): Parallel distributed processing (Explorations in the Microstructure of Cognition) - vol. 1: Foundations / vol. 2: Psychological and Biological Models. MIT Press, Cambridge (MA).

[19] Nagel, T. (Ed.) (1993): Experimental and Theoretical Studies of Consciousness. John Wiley & Sons, Chichester etc., 1993 (in particular: M. Kinsbourne: Integrated cortical field model of consciousness).

[20] Nelkin, N. (1996): Consciousness and the Origins of Thought. Cambridge Univ. Press, Cambridge.

[21] Newman, J. (1997): Toward a general theory of the neural correlates of consciousness. J. Consciousness Studies 4, 47-66 (part I) and 100-121 (II).

[22] Oakley, D.A. (Ed.) (1990): Brain and Mind. Methuen, London.

[23] Peretto, P. (1992): An Introduction to the Modeling of Neural Networks. Cambridge Univ. Press, Cambridge.

[24] Peruš, M. (1995a): All in One, One in All (Brain and Mind in Analysis and Synthesis). Ljubljana, DZS (in Slovene).

[25] Peruš, M. (1995b): Synergetic Approach to Cognition-Modeling with Neural Networks. In: K. Sachs-Hombach (Ed.): Bilder im Geiste. Rodopi, Amsterdam, Atlanta (183-194).

[26] Peruš, M. (1995c): Analogies between quantum and neural processing - consequences for cognitive science / In: P. Pylkkänen, P. Pylkkö (Eds.): New Directions in Cognitive Science. Finnish AI Soc., Helsinki (115-123).

[27] Peruš, M. (1996): Neuro-quantum parallelism in mind-brain and computers. Informatica 20, 173-183.

[28] Peruš, M. (1997a): Mind: neural computing plus quantum consciousnes. In: M. Gams, M. Paprzycki, X. Wu (Eds.): Mind Versus Computer. IOS Press, Amsterdam (156-170).

[29] Peruš, M. (1997b): Neuro-quantum coherence and consciousness. Noetic J. 1, in press.

[30] Peruš, M. (1997c): Common mathematical foundations of neural and quantum informatics. Z. Angewandte Mathematik und Mech., in press.

[31] Peruš, M. (1997d): System-theoretical backgrounds of mystical and meditational experiences. World Futures: J. General Evolution, in press.

[32] Peruš, M. (1997e): System-processual backgrounds of consciousness. Informatica 21, 491-506.

[33] Pribram, K.H. (1991): Brain and Perception. Lawrence Erlbaum, Hillsdale (NJ).

[34] Raković, D. & Dj. Koruga (Eds.) (1996): Consciousness. ECPD, Beograd.

[35] Sajama, S.; M. Kamppinen & S. Vihjanen (1987): A Historical Introduction to Phenomenology. Croom Helm, London etc.

[36] Searle, J.R. (1993): The problem of consciousness. Cognition & Consciousness 2, 310-.

[37] Tucson II (1996): Toward a Science of Consciousness. Consciousness Research Abstracts (JCS).

[38] Železnikar, A.P. (1990): On the Way to Information. Slovene Soc. Informatika, Ljubljana.

Kevin Kelly—An Interview

By *Anton P. Železnikar*

The reason for this interview arose through the reading of Kevin Kelly's book *Out of Control* on the Internet. Some author's views in this book come close to the philosophy and science of the informational developed in the last 12 years by the interviewer (the undersigned). The most significant concepts in the book concern emerging, evolution, and hive strategy of living beings. Although, as an extremely busy writer, editor, and journalist explorer, Kevin Kelly agreed for a short interview in five questions.

Who is Kevin Kelly for the readers of Informatica?

Kevin Kelly is the executive editor of **Wired**, Internet guru, journalist and writer in the field of telecommunication, net economy, and high-tech manufacturing. In 1994, he published the book *Out of Control*, the new biology of machines. In 1997, his new economy rules appeared in Wired as a new approach to the wired global economy.

Question 1 (Wed, 26 Nov 1997)

In the mainstream philosophy and science concerning AI, computer science, mathematics, etc. the concept (word) "emerging" ("of something") has a nonscientific, unrecognized, illegal, scince-fiction value. On the level of human experience "emerging" (also as being, informing) comes up as necessity in the material and living. For instance, consciousness of something emerges out of information lumps, changes, and vanishes.

The question is: *Emerging is the central concept of the informational (informing, counterinforming, embedding). Formal theories of chaos and randomness root in algorithms (mathematical, programmed sructures). A global web seems to possess a proper chaotic component in informational behavior of individual web users and in arising of the system hardware, both in an unforeseeable way.*

Does any constructive possibility for an implementation of information emerging in a spontaneous way exist, also within the future imaginable possibilities?

KK (Mon, 1 Dec 97):

As you suggest, the word "emerging" has an everyday meaning to us that does not hold up under scientific precision. We could just as well say "happens." Order happens from chaos. Nonetheless, we have hints that the everyday use of the word is correct sometimes, in certain conditions. Most clearly I think in evolutionary systems, where large sustainable patterns "emerge" or happen out of random, chaotic signals. What we don't have yet is rigorous mathematical equations to mirror

our everyday understanding.

APŽ (Wed, 3 Dec 1997):

Thank you for the answer. What I really agree upon is not only the insufficiency but also the deficiency of the scientific and formalistic (mathematical or mathetical—from the Greek $\mu\alpha\vartheta\eta\sigma\iota\varsigma$) concept (as you say, precision) of emerging. Emerging, happening, informing, and being are synonymous in the phenomenal sense. In the theory of the informational, operands and operators are emerging entities. This can easily and evidently be understood in the process of reading a text, especially the most meaningly difficult and causally circular one. (Footnote: I wrote a paper entitled "Informational Investigations", to appear this month in German, showing the circular graph and possible sentences generation based on eight sentences from the Heidegger's "Being and Time", treating understanding and interpretation.)[1] The new formalism already functions, but in the background of it the question of the constructive nature of emerging lurks, waiting for its formal implementation.

Question 2

Today, WWW (World-wide Web, Internet) is the most complex, spontaneously emerging (in user participation, hardware, and software), circularly perplexed informational system. Terminal points and nodes of WWW (users, human system operators) possess individual consciousness. This can be seen as a distributed consciousness in respect to a common system intention in communication among users, nodes, information sources, etc., however also in their informational improvement, change, disappearance. Now the real and substantial question is coming to the surface:

What and where is the consciousness of the system—WWW? How can it be grasped to bridge the gap between the understanding of human consciousness and the system consciousness? Not only in the sense of the so-called collective consciousness as an average in mass communication, category, characteristic of something, or the like. What could be comprehended as a system individual consciousness?

KK (Fri, 5 Dec 97):

That is the problem. Neurons cannot—in principle—comprehend thoughts of the brain. We are constitutionally—in principle—not going to be able to comprehend the emergent consciousness of the web.

[1] ŽELEZNIKAR, A.P. 1997. Informationelle Untersuchungen. Grunglagenstudien aus Kybernetik und Geisteswissenschaft/Humankybernetik 38:147–158.

APŽ (Sun, 7 Dec 1997):

A Comment to the Answer

Evidently, for made systems to become conscious, the complexity and individual consciousness (their conscious components) is necessary but not sufficient condition. This situation reminds on swarm consciousness (individual intention instead of the common system consciousness). A minimal form of organizational invariance must exist (e.g., a basic, principled, ruled organization), connecting the conscious components specifically in an informational sense, that is, consciously. In case of a social group (hive, swarm, mass), individuals are consciously or unconsciously coupled (genetically, instictivelly, intentionally). The conclusion is:

Besides system complexity and conscious individuality, a specific basic organization of the system is needed— the conscious one.

This kind of organization isn't explored yet, but it is presummed that its nature is circularly spontaneous, enabling the emerging of organization by the intrinsic and extrinsic informational impacts (signals). Probably, Internet could become artificially conscious only through the development and implementation of specific components, being distributed and connected in a conscious way, thus, cognizing the needs and problems of the Internet users. In this context, smart chaotic and random components of the system could be helpful too.

Question 3

In the index of your book "Out of Control" I didn't find the term "meaning" as a component of the use of signals (information, data) in made and social context. What is the main message of the "New Rules for the New Economy"? At the end of the paper (Wired: Sep 1997) I would expect a conclusion (also non-statistical experience), for instance, for managers entering the new domain of behavior and strategic thinking.

KK (Mon, 8 Dec 97):

Well, I had only room for an article in the magazine. I am now writing a short book that will try to address the question of what managers should do, what you call the "meaning" of it.

APŽ (Thu, 11 Dec 1997):

In your book *OoC,* evolution is a magic word behind which a human (linguistic) concept is hidden. Evolution is emerging of physically coupled entities in an unforeseeable way. Has it a designer or are entities of the System evolutionary designers by themselves?

Design of entities comes up as composition (out of "nothing", lumps, particles, energy quanta, etc.—

bottom-up design) and as decomposition (conception of structure and organization—top-down design).

Since the world is already here, the decomposition bothers me in a particular way—how to copy[2] the existing nature of entities. And as you say, direction and goals (of composition and decomposition) can emerge in biological evolution from a mob of directionless and goal-less parts (self-produced trend). Fine, I agree.

Question 4

Is consciousness of a constructor (composer and decomposer) evolutionary in every respect (out of control) in the sense that intention (trend) underlies the "rules" of evolution by itself?

KK (Mon, 22 Dec 97):

I am not entirely clear I understand your question.

I think that things that evolve can have either a direct conscious creator or not. Both work. For those that don't have an explicit creator, their evolution is governed by the implicit rules of the creator of the universe. We don't have very detailed understanding of what those indirect rules are, but my book was an attempt to describe what they may be like. These may be thought of as "embedded biases" in the universe. And of course, the rules may not be stable, but may evolve themselves according to meta-rules we don't understand.

APŽ (Thu, 8 Jan 1998):

I read intensively your book *OoC.* The question which came up very clearly is that between the postmodern American and traditional European philosophy the differences concerning nowadays global significance (importance, relevance) are substantial, in giving accent to different philosophical subjects.

Let me first paraphrase the dictum in your book, p. 567.

"Postmodern humans inform [swim] in a third transparent medium now materializing. Every fact that informs [can be digitized], is. Every measurement of collective human activity that can inform [be ported] over an information system [network], is. Every trace of an individual's life that can be informingly presented [transmuted into a number] and sent over web [a wire], is. This informing-in-the-world [wired planet] becomes a torrent of circular informing [bits circulating] in a clear shell of informing [glass fibers, databases, and input devices]. Once informing [moving], information [data] informs [creates] transparency. Once informed [wired], a society can understand (see) itself."

---

[2]It seem that APŽ (the undersigned) originally was thinking the "cope with" instead of the "copy". However, "to copy" possesses also an adequate meaning, that is the reflexive "to model oneself on copy".

The first paragraph of the dictum could belong to the Heideggerian phenomenology, if informing is replaced by being, and vice versa. At that time (Being and Time, 1927), informing of information was deeply under the philosophical horizon.

Your book, in fact, exposes a new look onto the possibilities of a postmodern philosophy—also philosophically. This philosophy emerges bottom-uply from research and technology projects, postmodern management, and human cognition in general. It does not come from a cathedra of a professional philosopher, his and his colleagues hive (also a kind of stub, stump, or stock), but spontaneously from the everyday experience of man (swarm logic and swarm experience).

Question 5

How to establish the new philosophy to become a practical (useful, evolutionary) guide in the postmodern world? Is its actual spreading over the net possible?

KK (Mon, 12 Jan 98):

I think it is possible to spread this philosophy over the web. We already see a converging sense of politics and music and science. I think it will continue to meld.

KK's Net and Book Information

Kevin Kelly
Wired magazine
520 3rd Street, San Francisco, CA 94107 USA
+1–415–276–5211 vox
+1–415–276–5150 fax
kevin@wired.com
www.hotwired.com/staff/kevin

The full text of his book OUT OF CONTROL can be found at
   http://www.well.com/user/kk/OutOfControl/

His next book on the NEW RULES OF THE NEW ECONOMY will be published in the fall of 98 by Viking/Penguin.

Out of "Out of Control"

Still available at Cox and Wyman Ltd., Reading, Berks, Great Britain, 666 pp.

Although pulled out of the context of OoC, dictums sound reasonable. Let us make a random choice of sentences in the book.

Tautology is, in fact, an essential ingredient of stable systems[506]. Life and evolution entail the necessary strange loop of circular causality—of being tautological

at a fundamental level[507]. Gödel's 1931 theorem demonstrates, among other things, that attempts to banish self-swallowing loopiness are fruitless, because, in Hofstadter's words, "it can be hard to figure out just where self-referencing is occurring[507]."

In 1991, a young Italian scientist, Walter Fontana, showed mathematically that a linear sequence of function A producing function B producing function C could be very easily circled around and closed in a cyberentic way into a self-generating loop, so that the last function was coproducer of the initial function[507]. ... *every* producer in the circuit is a producer of another, until every loop is incorporated into all other loops in massively parallel interdependence[508]. One small invention (the transistor) produces other inventions (the computer) which in turn permit yet other inventions (virtual reality)[509]. Where ideas are free to flow and generate new ideas, the political organization will eventually head toward democracy as an unavoidable self-organizing strong attractor[509]. "Order for free" flies in the face of a conservative science that has rejected every past theory of creative order hidden in the universe[511].

But wouldn't it be wonderful if somehow there are laws that make laws, so that the universe is, in John Wheeler's words, something that is looking in at itself? The universe posts its own rules and emerges out of a self-consistent thing[512].

In some way he[3] hoped to discover, evolutionary systems controlled their own structure. ... He was not content to show that order emerged spontaneously and inevitably. He also felt that *control* of that order emerged spontaneously.[512]

For a concise reader the following contents list may be of particular interest: The made and the born, Hive minds, Machines with an attitude, Assembling complexity, Coevolution, The natural flux, Emergence of control, Closed systems, Pop goes the biosphere, Industrial ecology, Networks economics, E-money, God games, In the library of form, Artificial evolution, The future of control, An open universe, The structure og organized change, Postdarwinism, The butterfly sleeps, Rising flow, Prediction machinery; Wholes, holes, and spaces; The nine laws of god (Distribute being, Control from the bottom up, Cultivate increasing return, Grow by chunking, Maximize the fringes, Honor your errors; Pursue no optima, have multiple goals; Seek persistent disequillibrium, and, Change changes itself).

A.P. Železnikar

---

[3]KAUFFMAN, S.A. August 1991. Scientific American.

KAUFFMAN, S.A. 1991. The Sciences of Complexity and 'Origins of Order'. Technical Report 91–04–021. Santa Fe Institute.

KAUFFMAN, S.A. 1993. The Origins of Order: Self Organization and Selection in Evolution. Oxford University Press.

Report

The 4th World Congress on Expert Systems

March 16–20, 1998, Mexico City

At the 4th World Congress on Expert Systems there were over 500 participants from 32 countries. The congress took place in ITESM, Mexico City on March 16-20, 1998. Overall, over 250 submissions were received, and 130 of them were included in the proceedings. There were 10 tutorials, 2 panels, and 6 workshops including Intelligent agents on the Internet. These events were performed mainly on March 16, the main conference started with invited lectures on March 17.

The intelligent agents workshop started with an interesting overview by Prof. Murugesan. Agents are a necessity for humans because humans can not control information explosion emerging on the Internet without appropriate tools. Terms like "infostress" or "information overkill" are replacing "infosphere". It is estimated that by 2000 most applications and information products will have agents integrated into them. There are different types of agents, e.g., agents with emotions that imitate human performance by facial expressions. There are shopping agents, service agents, interface agents, recommender agents, electronic commerce agents, information filtering agents etc. Like in any new discipline, agents are after a couple of years becoming more and more engineering oriented. SW engineering approach means that we want to produce systems with favourable cost/benefit ratio in a disciplined, systematic and controlled way. Agents, like any other SW product, have life cycles, need user manuals and other documentation. At the same time, agent research is moving towards new functions, e.g. social agents, agent protocols like KQML, negotiating agents etc. But an important property of agents remains to be orientation towards useful systems that perform human-assistance tasks on the Internet and computers in general. Agents are currently the most important SW computing paradigm, now and for at least the next decade.

The main conference events started on March 20 by invited lectures by E. A. Feigenbaum, L. A. Zadeh, A. Guzman-Arenas, D. Michie and others. Prof. Feigenbaum described his three years as Chief Scientist of the US Air Force from 1994 to 1997. In his words, there was world-class technology equipment derived from physical sciences, while information technologies were not always superb, sometimes even lagging advanced commercial products. All the services of the US Department of Defences, i.e. Army, Navy, and Air Force, use different artificial intelligence techniques in several applications. Among AI techniques, expert systems, intelligent agents and intelligent systems are often used.

Prof. Feigenbaum was an honorary conference chair. He delivered the fourth Feigenbaum Medal to Prof. Lafti A. Zadeh, the "father of fuzzy logic". Born in Baku, Azerbayean in 1921, he finished Ph.D. at the Columbia University. He received several prestigious awards, and is a member of the National Academy of Engineering. Prof. Zadeh is probably the most often cited AI scientists in SI and other major indexing databases.

In his presentation, Prof. Zadeh highlighted that humans perform a wide variety of tasks without any computations. A typical example would be parking a car. Humans employ their ability to perform tractable, robust, and effective. According to Prof. Zadeh, this essential human ability is closely linked to the modality of information granulation. In technical terms, he introduced granulation into fuzzy reasoning as two techniques: granular computing (GrC) and computing with words (CW). Both are well-defined theories built on mathematical foundations, capable of dealing with imprecision, uncertainty and partial truth. Both methods should enhance implementation of computer systems for real-life problems.

Prof. Michie presented recent findings in learning laws from data.

Prof. Jae Kyu Lee described new AI opportunities in electronic commerce. He emphasised the importance of AI applications, especially intelligent agents in future electronic commerce.

The author of this report presented a normal paper about the Slovenian intelligent employment agent, the first in the world to offer over 90% of all job possibilities.

The proceedings was edited by Francisco Cantu, Rogelio Soto, Jay Liebowitz, and Enrique Sucar. They did a great job gathering the most interesting presentations in the area of expert systems. In recent years, the basic slogan has been modified into: Expert Systems, Applications of Advanced Information Technologies. Among new approaches, electronic commerce, AI applications on the Internet and especially intelligent agents have progressed mostly in recent years.

Conference chair, Prof. Cantu-Ortiz and his team; Prof. Jay Liebowitz, the conference founder; and Honorary Conference Chair Prof. Feigenbaum organised an excellent event, the top world conference in expert systems. The next conference will be held in USA in 2000.

Matjaz Gams

Call for Paper
International Multi-Conference

Information Society - IS'98

6 – 9 October, 1998
Slovenian Science Festival
Cankarjev dom, Ljubljana, Slovenia

Programme committee:

dr. Cene Bavec, chairperson,
prof. dr. Ivan Bratko, co-chair,
prof. dr. Matjaž Gams, co-chair,
prof. dr. Tadej Bajd,
mag. Jaroslav Berce,
dr. Dušan Caf,
prof. dr. Saša Divjak,
dr. Tomaž Erjavec,
prof. dr. Nikola Guid,
prof. dr. Borka Jerman Blažič Džonova,
doc. dr. Gorazd Kandus,
doc. dr. Marjan Krisper,
mag. Andrej Kuščer,
prof. dr. Jadran Lenarčič,
dr. Franc Novak,
prof. dr. Marjan Pivka,
prof. dr. Vladislav Rajkovič,
prof. dr. Ivan Rozman,
dr. Niko Schlamberger,
prof. dr. Franc Solina,
prof. dr. Stanko Strmčnik,
prof. dr. Jurij Tasič,
prof. dr. Andrej Ule,
dr. Tanja Urbančič,
prof. dr. Baldomir Zajc,
dr. Blaž Zupan

Invitation

You are kindly invited to participate in the "New Information Society - (IS'98)" multi-conference to be held in Ljubljana, Slovenia, Europe, from 6-9 October, 1998. The multi-conference will consist of seven carefully selected conferences.

Basic information

The concepts of information society, information era, infosphere and infostress have by now been widely accepted. But, what does it really mean for societies, sciences, technology, education, governments, our lives? What are current and future trends? How should we adopt and change to succeed in the new world?

IS'98 will serve as a forum for the world-wide and national community to explore further directions, business opportunities, governmental European and American policies. The main objective is the exchange of ideas and developing visions for the future of information society. IS'98 is a standard high-quality scientific conference covering major recent achievements. Besides, it will provide maximum exchange of ideas in discussions, and concrete proposals in final reports of each conference.

The multi-conference will he held in Slovenia, a small European country bordering Italy and Austria. It is a land of thousand natural beauties from the Adriatic sea to high mountains. In addition, its Central European position enables visits to most European countries in a radius of just a few hours drive by car. The social programme will include trips by desire and organised trips to Skocjan or Postojna caves. Coffee breaks, the conference cocktail and dinner will contribute to a nice working atmosphere.

Call for Papers

Deadline for paper submission: 15 June, 1998

Registration fee is 100 US $ for regular participants (6.000 SIT for participants from Slovenia) and 50 US $ for students (3.500 SIT for Slovenian students). The fee covers conference materials and refreshments during coffee-breaks.

More information

For more information visit
`http://turing.ijs.si/is/indexa.html` or contact
`milica.remetic@ijs.si`.

The multi-conference consists of the following conferences:

Information Society

6–7 October, 1998
Chairs: dr. Cene Bavec, prof. dr. Matjaž Gams
Contact person: prof. dr. Matjaž Gams
Phone: (+386 61) 1773 644
E-mail: `matjaz.gams@ijs.si`
Jozef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia, Europe

Language Technologies

Date: 6–7 October, 1998
Chair: dr. Tomaž Erjavec
Contact person: dr. Tomaž Erjavec
Phone: (+386 61) 1773 644
E-mail: `tomaz.erjavec@ijs.si`
Address: Jozef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia, Europe
Submission deadline: 15 June, 1998

Manufacturing Systems and Technologies

Date: 7 October, 1998
Chair: prof. dr. Jadran Lenarčič
Contact person: prof. dr. Jadran Lenarčič
Phone: (+386 61) 1773 378
E-mail: `jadran.lenarcic@ijs.si`
Address: Jozef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia, Europe
Submission deadline: 31 August, 1998

Education and Information Society

Date: 8 October, 1998
Chair: prof. dr. Vladislav Rajkovič
Contact person: Mojca Florjančič
Phone: (+386 064) 22 10 61
E-mail: `mojca.florjancic@fov.uni-mb.si`
Address: Faculty of Organizational Sciences, Kidričeva 55a, 4000 Kranj, Slovenia, Europe
Submission deadline: 15 June, 1998

Development and Reingeneering of Informaton Systems

Date: 8 October, 1998
Chair: prof. dr. Ivan Rozman
Contact person: dr. Ivan Rozman
Phone: (386 62) 2207 410
E-mail: `i.rozman@uni-mb.si`
Address: FERI, Smetanova 17, 2000 Maribor, Slovenia, Europe
Submission deadline: 15 June, 1998

Cognitive Sciences

Date: 9 October, 1998
Chair: prof. dr. Andrej Ule
Contact person: prof. dr. Andrej Ule
Phone: (061) 1769 200

E-mail: `andrej.ule@guest.arnes.si`
Address: Jozef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia, Europe
Submission deadline: 15 June, 1998

Computer Analysis of Medical Data

Date: 9 oktober 1998
Chair: dr. Blaž Zupan
Contact person: dr. Blaž Zupan
Phone: (061) 1773 380
E-mail: `blaz.zupan@ijs.si`
Address: Jozef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia, Europe
Submission deadline: 15 June, 1998

CC AI

The journal for the integrated study
of Artificial Intelligence, Cognitive Science
and Applied Epistemology.

CC-AI publishes articles and book reviews relating to the evolving principles and techniques of Artificial Intelligence as enriched by research in such fields as mathematics, linquistics, logic, epistemology, the cognitive sciences and biology.
CC-AI is also concerned with development in the areas of hard- and software and their applications within AI.

Editorial Board and Subscriptions

CC-AI, Blandijnberg 2, B-9000 Ghent, Belgium.
Tel.: (32) (9) 264.39.52,
Telex RUGENT 12.754
Telefax: (32) (9) 264.41.97
e-mail: `Carine.Vanbelleghem@RUG.AC.BE`

Call for Papers

Advances in the Theory and Practice of Natural Language Processing

Special Issue of Informatica 22 (1998) No. 4

Informatica, an International Journal for Computing and Informatics, announces the Call for Papers for the issue of an interdisciplinary volume dedicated to the theoretical and practical aspects of natural language (NL) analysis and generation.

This special issue is intended to be a forum for presenting, first of all, the theoretical ideas proved to be effective in the design of NL processing systems or promising to considerably extend the sphere of successful NLP applications.

TOPICS: Original papers are invited in all subareas and on all aspects of NLP, especially on:

1. The current state and advancements in the last five years in particular subfields of NLP.

2. Natural-language-like knowledge and meaning representation formal systems.

3. Formal approaches to describing conceptual structures of complicated real discourses (pertaining, e.g., to medicine, technology, law, business, etc.).

4. New logics for NLP.

5. Semantics-oriented methods of natural language analysis, conceptual information retrieval in textual data bases.

6. Computational lexical semantics, ontologies for NLP.

7. Understanding of metaphors and metonymy.

8. Anaphora resolution.

9. Generation of natural language discourses.

10. Parallel conceptual processing of natural language texts.

11. Intelligent text summarization.

12. New directions in NLP.

Informatica 22 (1998) No. 4, in an enlarged volume, is fixed as the special issue.

Time Table and Contacts

The deadline for the paper submission in four copies is May 30, 1998.

Printed-paper mail address:
Prof. A.P.Zeleznikar, Jozef Stefan Institute, Jamova c. 39, SI-1111 Ljubljana, Slovenia.

Correspondence (e-mail addresses):

— anton.p.zeleznikar@ijs.si
    Prof. Anton P. Železnikar, Slovenia
— vaf@nw.math.msu.su
    Prof. Vladimir A. Fomichov, Russia
— kitano@csl.sony.co.jp
    Prof. Hiroaki Kitano, Japan

Format and Reviewing Process

As a rule, papers should not exceed 8,000 words (including figures and tables but excluding references. A full page figure should be counted as 500 words).
Ideally 5,000 words are desirable.

Each paper will be reviewed by at least two anonymous referees outside the author's country and by the appropriate editors.

In case a paper is accepted, its author (authors) will be asked to transform the manuscript into the Informatica LaTeX style (available from ftp.arnes.si; directory: /magazines/informatica).

For more information about the Informatica and the Special Issue see
    FTP: ftp.arnes.si
with anonymous login or
    URL:
    http://turing.ijs.si/Mezi/informat.htm.

Call for Contributions and Session Proposals

International Conference on

CONSCIOUSNESS IN SCIENCE AND PHILOSOPHY '98
November 6–7, 1998, Charleston, Illinois, U.S.A.

Conference Homepage: http://www.ux1.eiu.edu/~cfskd/conference.htm

Eastern Illinois University, Department of Mathematics, is the host organizer of the 1st International Conference on Consciousness in Science and Philosophy. The scientific program of the conference will include the following sessions (probably about 2-3 hours each, 20-30 minutes, exceptionally, 60 minutes per lecture) organized by chairmen whose names are written after each topic, and any new chairmen for new suggested topics:

1. Informational phenomenalism of consciousness
   Prof. Anton P. Železnikar,
   anton.p.zeleznikar@ijs.si for long texts,
   and at home s51em@lea.hamradio.si;
   indexed home page with different documents at
   URL http://lea.hamradio.si/~s51em/

2. Cybernetic concepts of consciousness
   Prof. Jerry L.R. Chandler,
   jlrchand@erols.com

3. Cognitive aspects of consciousness
   Prof. Vladimir A. Fomichov,
   vaf@nw.math.msu.su

4. Complex system background of consciousness
   Mitja Perus,
   mitja.perus@uni-lj.si

5. Physical background of consciousness
   Prof. Richard Amoroso,
   ramoroso@hooked.net

6. Consciousness and computers
   Dr. Ben Goertzel,
   ben@goertzel.org

7. Transcendental states of consciousness
   Prof. Suhrit K. Dey,
   cfskd@eiu.edu

8. Practical application of consciousness-based technology
   Dr. Ken Walton,
   kwalton@mum.edu

9. Conception of minds as semiotic systems
   Prof. James H. Fetzer,
   jfetzer@d.umn.edu

10. Neural and psychological correlates of consciousness
    Dr. Fred Travis

11. Biosystem aspects of consciousness
    Prof. Igor Ackchurin & Dr. Serge Konyaev
    skonyaev@iphras.irex.ru,
    better use mitja.perus@uni-lj.si

Additionally to these sessions, new sessions can be suggested. *Any potential new chairman willing to organize an additional session, please, send proposed session title and a description of the topic as soon as possible (deadline May 1, 1998) to* mitja.perus@uni-lj.si .

The following session topics, for example, are open:

– Philosophical/psychological aspects of consciousness

– Neural correlates of consciousness

– Models of natural and artificial consciousness

etc.

Other consciousness studies topics are welcome also.

Deadline for Submition of Summaries

Please, send your lecture summary by e-mail to the chairman of the session where you would like to take part and to mitja.perus@uni-lj.si . If you feel that your contribution does not fit well into any of the mentioned sessions, please send your summary by e-mail to mitja.perus@uni-lj.si and add your classification suggestion (e.g., neural correlates of consciousness, psychological aspects, models, etc.)

Deadline for summaries is May 15, 1998.

All correspondence will be communicated by e-mail.

Host chairman and local organizer:
Prof. S.K. Dey
Eastern Illinois University
Department of Mathematics, 600 Lincoln Avenue
Charleston, Illinois, 61920–3099 USA
e-mail: cfskd@eiu.edu

Program secretary:

Mitja Perus
National Institute of Chemistry
Hajdrihova 19 (POB 3430)
SI–1001 Ljubljana, Slovenia
& Slovene Society for Cognitive Sciences
e-mail: mitja.perus@uni-lj.si

It is expected that all chairmen will send the conference calls and information to their session lecturers and other potential attendees around the world—by e-mail or by ordinary mail. They are requested to send session program information, i.e. all names of lecturers and their titles, and suggestions to mitja.perus@uni-lj.si *as soon as possible, but the deadline for session chairmen is mAY 15, 1998.*

Notification of acceptance of lectures by e-mail due July 1, 1998.

Conference home page, made by Dr. Dheeraj Bhardwaj,
http://www.ux1.eiu.edu/~cfskd/conference.htmis in preparation. It will include all information about acommodation and about conference venue at the Eastern Illinois University, including registration form (registration form already there). Please, follow the homepage updates.

Main information sources (for those not having access to WWW):
cfskd@eiu.edu (organization, facilities)
mitja.perus@uni-lj.si (program).

Registration fee: $ 150

PROCEEDINGS of the conference will be published as a special issue of the *Journal of Applied Science & Computation* or of the journal *Informatica.*

Accomodation information

There are TWO VERY nice motels:

1. BEST WESTERN (Worthington Inn)

    – Single Room $46.00
    – Double Room $46.00
    – King Suite $55.00

   Phone: 217 348 8161
   FAX : 217 348 8165
   Address: 920, West Lincoln, Charleston, IL 61920
   USA
   Attention: Ms. Kara Smith

2. ECONO-LODGE

    – Single Room $42
    – Double Room $46
    – King Size $49

Phone: 217 345 7689
FAX : 217 345 7697
Address: 810 West Lincoln, Charleston, IL 61920
USA
Attention: Mr.Rakhes Patel

All the participants have to make their **own** room reservation as soon as possible. Because during that time there are too many **sports** events in the town and the motels **do not** want to keep any room unreserved.

All these motels accept Visa & Mastercard.

3rd International Conference on

COMPUTATIONAL INTELLIGENCE AND NEUROSCIENCE
Sheraton Imperial Hotel & Convention Center
Research Triangle Park, North Carolina
October 24–28, 1998 (Tutorials are on October 23)

Conference Co-chairs

Subhash C. Kak, Louisiana State University
Jeffrey P. Sutton, Harvard University

This conference is part of the Fourth Joint Conference Information Sciences.

Organizing Committee

- George M. Georgiou, California State University, Chair

- Grigorios Antoniou, Griffith University, Australia

- Catalin Buiu, Romania

- Ian Cresswell, U.K.

- S. Das, University of California, Berkeley

- S.C. Dutta Roy, India

- Laurene Fausett, Florida Institute of Technology

- Paulo Gaudiano, Boston, University

- Masafumi Hagiwara, Keio Univ., Japan

- Ugur Halici, METU, Turkey

- Akira Hirose, University of Tokyo

- Arun Jagota, University of North Texas

- E.V. Krishnamurthy, Australian National University, Canberra

- Ping Liang, University of California, Riverside

- Jacek Mandziuk, Warsaw Univ. of Technology, Poland

- Jonathan Marshall, University of N. Carolina

- Bimal Mathur, Rockwell CA

- Kishan Mehrotra, Syracuse

- Ouri Monchi, King's College London, UK

- Haluk Ogmen, University of Houston

- Ed Page, South Carolina

- Mitja Perus, National Institute of Chemistry, Slovenia

- Vladimir Radevski, Univ. of Paris, FRANCE

- Raghu Raghavan, National Univ. of Singapore

- W.A. Porter, University of Alabama

- Ed Rietman, Bell Labs

- Christos Schizas, University of Cyprus

- Harold Szu, USL

- M. Trivedi, UCSD

- Nicolae Varachiu, National Institute of Microtechnology, Romania

- E. Vityaev, Russia

- Paul Wang, Duke University

- Sumio Watanabe, Gifu Univ., Japan

- Edward K. Wong, Polytechnic University, NY, USA

Plenary Speakers include the following:

James Anderson
Panos J. Antsaklis
John Baillieul
Walter Freeman
David Fogel
Stephen Grossberg
Yu Chi Ho
Thomas S. Huang
George J. Klir
Teuvo Kohonen
John Koza
Richard G. Palmer
Zdzislaw Pawlak
Azriel Rosenfeld
Julius T. Tou
I.Burhan Turksen
Paul J. Werbos
A.K.C. Wong
Lotfi A. Zadeh

Hans J. Zimmermann
Karl Pribram
Stuart Hameroff

Areas for which papers are sought include:

Artificial Life
Artificially Intelligent NNs
Associative Memory
Cognitive Science
Computational Intelligence
Efficiency/Robustness Comparisons
Evolutionary Computation for Neural Networks
Feature Extraction & Pattern Recognition
Implementations (electronic, Optical, Biochips)
Intelligent Control
Learning and Memory
Neural Network Architectures
Neurocognition
Neurodynamics
Optimization
Parallel Computer Applications
Theory of Evolutionary Computation

Summary Submission Deadline: June 1, 1998

Decision & Notification: August 1, 1998

Papers will be accepted based on summaries. A summary shall not exceed 4 pages of 10-point font, double-column, single-spaced text, (1 page minimum) with figures and tables included. Send 3 copies of summaries to:

George M. Georgiou
Computer Science Department
California State University
San Bernardino, CA 92407–2397
U.S.A.
georgiou@csci.csusb.edu

CONFERENCE REGISTRATION FORM, PUBLICATION INFORMATION, tutorial and other registration information can be found in the announcement of the Fourth Joint Conference Information Sciences on the conference Web site:
    http://www.csci.csusb.edu/iccin

There will be a special session entitled NEURO-QUANTUM INFORMATION PROCESSING (information, registration, suggestions: mitja.perus@uni-lj.si)

Some TOPICS of the neuro-quantum session:

- neural networks and quantum computing

- brain as a multi-level information processing system incorporating neural, sub-cellular, quantum, sub-quantum collective dynamics

- mathematical and system-theoretical analogies in neural and quantum models

- informational biophysics and nanobiology

- neural and quantum holography

- neuro-quantum computers

- questions of biological plausibility of neural/quantum network models of brain processes

- neuro-quantum background of consciousness

- holographic/holonomic brain theory, Bohmian implicate order, Hopfield model and attractor networks

- Deutsch-like quantum computers

- quantum-inspired artificial neural net models

- complex system theory and emergent computation

- computational cognitive neuroscience and quantum biology

- modular compatibility of neural/quantum models

Invitations were send to the following distinguished neuro-quantum information processing experts:

CONFIRMED INVITED LECTURERS of the neuro-quantum session: Laurence Gould, Stuart Hameroff (plenary), Mari Jibu, Bruce MacLennan, Masanori Ohya, Karl Pribram (plenary), Giuseppe Vitiello, Kunio Yasue

INVITED LECTURERS NOT YET CONFIRMED: Robert Alicki, Friedrich Beck, Walter Schempp, Branko Soucek, Robert Turner (Also invited neural/quantum informatics experts: Manger, Marcer, Nobili, Scott, Sutherland, Ventura)

If you want to take part at THIS SESSION, please send two copies of summaries until June 1, 1998, marked by FOR NEURO-QUANTUM INFORMATION PROCESSING SESSION, also to one of the following addresses of Mitja Perus:

> Mɪᴛᴊᴀ Pᴇʀᴜs
> National Institute of Chemistry
> L01 / Lab for Molecular Modeling and NMR
> Hajdrihova 19 (POB 3430)
> SI–1001 Ljubljana; Slovenia
> E-mail: mitja.perus@uni-lj.si
> WWW: http://kihp6.ki.si/ mitja/index1.html
> Fax: (+386-61)-1259-244, (+386-61)-1257-069
> Tel: (+386-61)-1760-275, (+386-61)-1760-314
> Slovene Society for Cognitive Sciences

Second Conference Announcement & Call for Papers

INAUGURAL CONFERENCE FOR THE SOCIETY FOR THE MULTIDISCIPLINARY STUDY OF CONSCIOUSNESS
August 17 and 18, 1998
Fort Mason Center for the Arts
San Francisco, California

The Society for the Multidisciplinary Study of Consciousness is a network of people whose goals are to investigate, understand, and disseminate information concerning the topics of consciousness. These goals will be sought within the context of an ever-developing search for basic scientific concepts and underlying principles that are valid across many disciplines.

Conference Theme

Within the broad framework of cognitive science, the first meeting of the Society will focus upon ways we can encourage and organize the establishment of multidisciplinary conceptions and principles of consciousness. Toward this end submissions are invited which address topics of consciousness within the various disciplines and areas of cognitive science, including cognitive psychology, cognitive neuroscience, linguistics, philosophy of mind, artificial intelligence, and cognition and nonlinear dynamics.

Organizing Committee

- Brad Challis, University of Tsukuba, cognitive psychology

- Anne Jaap Jacobson, University of Houston, philosophy

- Earl Mac Cormac, Duke University, radiology

- Paavo Pylkkanen, University of Skoevde, philosophy

- Maxim Stamenov, Bulgarian Academy of Sciences, Linguistics

- Larry Vandervert, American Nonlinear Systems, cognitive neuroscience

- Philip Zelazo, U. of Toronto, cognitive & developmental psychology

Keynote Address: Karl H. Pribram, Director, Center for Brain Research and Informational Sciences, Radford University

"Conscious and Unconscious Processes: Relation to the Deep and Surface Structure of Memory"

Plenary Panel

Panel Title: METAPHORS OF CONSCIOUSNESS

The aim of this panel will be to discuss the relationship between consciousness and the representations of consciousness in the context of consciousness modeling in the cognitive sciences. The topics of potential interest include (but are not limited to) the following:

- What are the basic metaphors used in representing consciousness, e.g., the theater metaphor, and so forth;

- Are there preferences in different disciplines within cognitive sciences to prefer some metaphor(s) at the expense of others and what are the possible motivations for this state-of-affairs;

- What are the relationships between the phenomenon of consciousness and the images, concepts, models, and metaphors of it;

- What are the possibilities to develop integrative metaphors for consciousness representation and modeling?

Although the plenary panel is now set, persons wishing to present separate papers on the panel theme are encouraged to submit such abstracts to Larry Vandervert (see instructions below).

Plenary Panel Members

- Maxim I. Stamenov (Co-chair) Bulgarian Academy of Sciences; linguistics
  E-mail: maxstam@bgearn.acad.bg

- Earl R. Mac Cormac (Co-chair) Duke University; radiology

- Bernard Baars, Wright Institute, Berkeley; cognitive science

- Ralph Ellis, Clark Atlanta University; philosophy

- Charles Li, Univ. of California, Santa Barbara; linguistics

- Bruce Mangan, U. of California, Berkeley; cognitive science

Submission of Abstracts for Papers and Posters

Acceptance of submissions depends on their suitability to the CONFERENCE THEME, quality, and availability of slots. Any person may make only one oral presentation or poster presentation, but may be co-author on more than one. Oral presentations are limited to 25 minutes plus a 5 minute discussion period. Concurrent sessions will take place at adjacent sites at Fort Mason Center for the Arts.

Submissions must include ALL of the following information:

1. Title, Abstract (limited to one single-space page)

2. Name(s), Institutional Affiliation(s)

3. Postal Address(es), E-mail Address(es)

4. Telephone and FAX numbers

5. Indicate which co-author will present.

6. Indicate whether the presentation will be in spoken or poster form.

7. If a spoken presentation can not be fit into the available time slots, indicate your willingness to present in the poster session format.

Deadline for submission of abstracts is march 15, 1998. Send submissions to (regular mail preferred):

Dr. Larry R. Vandervert
American Nonlinear Systems
1529 W. Courtland
Spokane, WA 99205-2608 USA
Phone:+(509) 533-3583; FAX: +(509) 533-3149
E-mail: larryv@sfcc.spokane.cc.wa.us

CONFERENCE REGISTRATION FEE IS $145 (US funds).

The registration fee includes a copy of selected conference proceedings (approximately 350 pages) to be published by John Benjamins Publishing, all conference sessions, and a one-year membership in THE SO-CIETY FOR THE MULTIDISCIPLINARY STUDY OF CONSCIOUSNESS. Remit conference registration fee to Dr. Vandervert. Non-participant (including students) registration fee for those wishing a copy of the proceedings is $75., otherwise non-participating registration fee is $10.

Special travel and accommodation packages have been arranged through Carlson Wagonlit Travel, 1550 E.

Battlefield, Suite g2, Springfield, MO 65804 USA. Attention: Debra Nordberg. For more information call Debra Nordberg at +800-641-4331 or fax +417-883-5838.
E-mail: dnordberg@cwt-mcdanieltravel.com

Note: Because the American Psychological Association will be holding its national convention in San Francisco during the same week (August 14–18) with approximately 20,000 in attendance, we have secured these special accommodations near Fort Mason. But space is limited, therefore it is recommended that reservations be made early.

This conference is supported by

- THE CENTER FOR BRAIN AND INFORMA-TIONAL SCIENCES, Radford University, Radford, VA, USA and

- John Benjamins Publishing, publishers of the AD-VANCES IN CONSCIOUSNESS RESEARCH series.

Call for Papers

Working Group 8.3 of the International Federation for Information Processing invites you to participate in its 1998 working conference on

CONTEXT-SENSITIVE DECISION SUPPORT SYSTEMS

LOCATION: Bled, Slovenia

DATES: 13-15 July 1998 (opening reception on Sunday, July 12th)

ORGANIZED BY: IFIP Working Group 8.3 on Decision Support Systems; The University of Maribor, Faculty of Organizational Sciences and The Slovene Society Informatika

The focus of this conference is on issues related to developing context-sensitive decision support systems (DSS). There are a number of contexts that need to be taken into account (e.g., cultural, organizational, task-, role- or individual-related), depending on the purpose of the DSS and its target user(s). These contexts interact with and influence each other and an appreciation of their importance, in their totality, in design decisions can give rise to DSS which are adaptable to different environments and circumstances.

The adaptability of a decision support system should be considered along two different dimensions. On one hand, we have the horizontal dimension which considers changes though time within a particular context. That is, organizations and their practices change their requirements for decision support and technology continuously progresses to the effect that designers are given a much wider range of technical possibilities for providing decision support in a manner that could be more effective. Moreover, specific DSS face the problem where data, information and knowledge are continuously evolving. On the other hand, we have the vertical dimension of adaptability which relates to the transferability of the DSS to different (cultural or organizational) contexts. DSS, designed on the basis of an image of a generalizable task or role, irrespective of its context, will fail their purpose because they may be too general to support fully even the original intended user(s) because his/her cultural and organizational context has not been taken into account.

This conference aims to initiate a discussion on these issues which are of vital importance to designing effective and adaptable DSS. It will welcome contributions from all disciplines as the issues in question cut across different disciplines, each making its unique contribution to the topic.
The Conference Goal is:
How can we bring about a more useful, context-sensitive, generation of DSS?

The pivotal issues the conference wishes to address are:
How one can understand the context within which one designs and implements a DSS?
How one can model, represent and use context in a DSS?
How may context-sensitivity improve the effectiveness of DSS?

We invite papers and panel proposals related to:

- design of DSS for reasoning about context

- design of a DSS that takes into account its context

- interaction of choice and context

- impact of context on DSS success

- role of context in the decision maker-DSS interaction/cooperation

- role of context in the decision maker-DSS complementarity

- role of context in the knowledge organization of the DSS and its reasoning

- context and alternatives (selection, argumentation, explanation)

- context and decision criteria

- change and adaptability of DSS within one organization

SUBMISSION DATES:

- Full abstract (500-600 words):1 October 1997

- Deadline for submission of paper: 9 January 1998

- Notification of acceptance:13 February 1998

- Camera-ready copy due:16 March 1998

INSTRUCTIONS TO AUTHORS:

Only original unpublished papers should be submitted. All submissions will be reviewed. Selection for presentation at the conference and publication in the proceedings will be based on originality, contribution to the field, and relevance to the conference theme. The conference book, published by Chapman and Hall,

will be distributed at the conference and at least one author for each paper must register for the conference and present the paper.

Papers must not exceed 12-15 pages when single spaced. All submissions must include on the first page: title, author's name(s), affiliation, complete mailing address, phone number, fax number, and email address. An abstract of 100 words maximum and up to five keywords should be included before the body of the paper. Papers must be submitted in electronic form, using the Chapman and Hall Word template UKdoc.doc, which can be found on the web at http://www.it-ch.com/itch/authors/macros.html. This template contains detailed guidelines on how to format your paper. It is very important that you load the template and type within it using the automatic style guidelines it gives.

We would like to request that a 500-600 word abstract for your paper be submitted by 1 October 1997 for comments. Submissions can be sent by e-mail to: Dina Berkeley at the London School of Economics.

For further information, please contact program committee members:

George R. Widmeyer (Conference Chair)
University of Michigan Business School,
701 Tappan Street, Ann Arbor,
Michigan 48109-1234,USA
phone: +1 313 763 5808, fax:+1 313 764 3146
e-mail: widmeyer@umich.edu

Dina Berkeley
London School of Economics,
Dept. of Social Psychology, Houghton Str., London
WC2A 2AE, UK
phone: +44 171 9557401, fax: +44 171 9163864
e-mail: d.berkeley@lse.ac.uk

Patrick Brezillon
LAFORIA - UPMC - Case 169, 4 place Jussieu -
F-75252 Paris Cedex 05, France
phone: +33 1 44 27 70 08, fax: +33 1 44 27 70 00
e-mail: patrick.brezillon@lip6.fr

Vladislav Rajkovic (Organizing Chair)
Univerza v Mariboru,
Fakulteta za organizacijske vede,
Presernova 11, SI-4000 Kranj, Slovenia
phone: +386 61 1403301, fax: +386 61 1403301
e-mail: vladislav.rajkovic@ijs.si

Visit our web site at:
http://www-personal.umich.edu/
~widmeyer/ifipwg83

Machine Learning List

The Machine Learning List is moderated. Contributions should be relevant to the scientific study of machine learning. Mail contributions to ml@ics.uci.edu. Mail requests to be added or deleted to ml-request@ics.uci.edu. Back issues may be FTP'd from ics.uci.edu in pub/ml-list/V<X>/<N> or N.Z where X and N are the volume and number of the issue; ID: anonymous PASSWORD: <your mail address> URL-http://www.ics.uci.edu/AI/ML/Machine-Learning.html

First Call for Papers

International Conference on Systems, Signals, Control, Computers (Sscc'98)

International Association for the Advancement of Methods for System Analysis and Design (Iaamsad) and Academy of Nonlinear Sciences (Ans)

Announce the International Conference on
Systems, Signals, Control, Computers (Sscc'98)
Durban, South Africa (September 22-24, 1998)

and Invite Potential Authors for Submission of Papers

A preliminary WEB home page can be accessed on
    http://nsys.ntech.ac.za/iaamsad/
        SSCC98test.html
This home page will become public when International Programme Committee membership become confirmed.

Honorary Chairman: Academician V.M.Matrosov (Russia)
Conference Chairman: V.B.Bajic (South Africa)

Advisory Board:

V.B.Bajic (South Africa), J.Brzobohaty (Czech Republic), P.Daoutidis (USA), W.Hide (South Africa), C.Morabito (Italy), V.V.Kozlov (Russia), P.Leach (South Africa), P.C.Muller (Germany), L.Shaikhet (Ukraine), E.Rogers (UK), H.Szu (USA).

International Programme Committee:

V.Apanasovich (Belarus), V.B.Bajic (South Africa), C.Berger-Vachon (France), J.Brzobohaty (Czech Republic), M.Campolo (Italy), P.Daoutidis(USA), T.Fukuda(Japan), Z.Gajic (USA), M.Gams (Slovenia), J.Gil Aluja (Spain), Ly.T.Gruyitch (France), H.Hahn (Germany), M.Hajek (South Africa), R.Harley (South Africa), W.Hide (South Africa), M.Jamshidi (USA), V.Kecman (New Zealand), B.Kovacevic (Yugoslavia), V.Krasnoporoshin (Belarus), V.V.Kozlov (Russia), P.Leach (South Africa), L.K.Kuzmina (Russia), V.Milutinovic (Yugoslavia), C.Morabito (Italy), P.C.Muller (Germany), H.Nijmeijer (The Netherlands), D.H.Owens (UK), D.Petkov (South Africa), K.M.Przyluski (Poland), E.S.Pyatnitskii (Russia), E.Rogers (UK), L.Shaikhet (Ukraine), A.V.Savkin (Australia) H.Szu (USA),

E.I.Verriest (USA), R.Vrba (Czech Republic), J.Ziska (Czech Republic).

Local Organizing Committee:

V.Bajic, P.Govender, R.Hacking, M.Hajek, M.McLeod, K.S.Moodley, R.Papa, C.Radhakishun, A.Singh.

Address Of The Conference Office:

Sacan, P.O.Box 1428, Link Hills 3652, Durban, South Africa Tel./Fax: (+27 31) 204-2560 e-mail: bajic.v@umfolozi.ntech.ac.za

Supporting Organizations:

SANBI - South African National Institute for Bioinformatics (South Africa)
SAICSIT - South African Institute for Computer Scientists and Information Technologists (South Africa)
CER - Centre for Engineering Research, Technikon Natal (South Africa)
M L Sultan Technikon (South Africa)

General Information

1998 year is the year of Science and Technology in South Africa. The intention of the Department of Arts, Culture, Science and Technology of South Africa is to make South Africans more aware of how Science and Technology affects them in every-day life. Such a national initiative is in a way a very good environment for a conference like this: one that has a broad scope and spans many different fields. At the same time an opportunity is given to the research community of

South Africa to interact more directly with overseas peers.

Aims And Scope

The Conference is broad in scope and will provide a forum for the exchange of the latest research results as applied to different branches of science and technology. The areas of interest include concepts, techniques and paradigms associated with systems, signals, control and/or computers.

Domains of application include informatics, biomedical technology, economics, management, diverse engineering and science fields and applied mathematics. Artificial intelligence techniques are of particular interest, as well as reports on industrial applications.

The conference will include several plenary and invited lectures from world renowned scientists and regular papers. A number of special and invited sessions will also be organised, dealing with focussed areas of interest. The proposals for these special sessions should be submitted at the same time as the abstracts. A special session cannot have less than three papers or more than six.

The official language of the conference is English.

Manuscript Submission And Review Process

Three copies of the extended abstract (at least two pages) should be sent to the Conference Office at the address given below. Full papers are preferred. Papers in Microsoft Word can be sent by e-mail. All submissions will be reviewed by members of the International Programme Committee; additional reviewers will be consulted if necessary. The submissions will be reviewed as soon as they arrive; the average review time is about four weeks. Authors of accepted papers will thereafter be informed (by e-mail if available) of the required format for camera-ready paper submissions. In order for reviewers to be able to assess the submissions, the extended abstract has to provide sufficient information about the background to the problem, the novelty of the obtained results and the results achieved, the conclusions drawn and some references. Up to five keywords should be provided. All submitted papers have to be original, unpublished and not submitted for publication elsewhere.

Proceedings

All accepted papers will be published in the Conference Proceedings, which will be issued by a renowned international publisher.

Important Notice

Although we expect that the authors of accepted papers will present the papers at this Conference, we recognize that circumstances may prevent authors from participation at the Conference. In such cases the accepted papers will be published if the authors inform organizers of their non-attendance at the Conference by 15th May 1998. However, conference fees according to established rules have to be pre-paid in order that papers appear in the Proceedings.

Conference Fees

The conference fee for one participant covers the publication of two papers (each with a maximum of five A4 pages in length) according to the required format; one volume of the Proceedings in which the paper(s) appear(s); refreshment during the conference; one lunch and a banquet. Additional volumes of the Proceedings can be purchased for US$ 55.00. Authors of multiple papers are to pay additional fees for extra papers according to the specified rule. Social programme and tourist visits will be provided at extra cost.

Reduced registration fee of US$ 280.00 (South Africans R 1120.00) is applicable for early received, reviewed and accepted papers for which fee is paid by February 25, 1998 - prospective authors are encouraged to take advantige of this convenience; otherwise the following rates apply:

Early registration fee: US$ 350.00 (South Africans R 1400.00)

Late and on-site registration fee: US$ 400.00 (South Africans R 1600.00)

Student fee: US$ 200.00 (South Africans R 800.00) - to qualify for the student scale of fees, all authors mentioned on the paper have to be current students; written proof has to be provided at the time of payment

Payment in South African rands is possible only when all authors of the papers are South African residents; written proof has to be provided at the time of payment.

Deadlines

Extended Abstracts and Special Session Proposals:
- submission by mail (15th February, 1998)
- submissions by e-mail (15th January, 1998)
Notification of acceptance (15th April, 1998)
Submission of papers in camera ready form (15th May, 1998)
Early payment of conference fees (15th May, 1998)
Late payment of conference fees (31 June, 1998)

First Announcement and Call for Papers
JCKBSE'98
Third Joint Conference on
Knowledge-Based Software Engineering
Smolenice, Slovakia,
September 9-11, 1998

Sponsored by:

SIG on Knowledge-Based Software Engineering, Institute of Electronics, Information and Communication Engineers (EICE), Japan
Faculty of Electrical Engineering and Information Technology, Slovak University of Technology in Bratislava, Slovakia
Russian Association for Artificial Intelligence
Bulgarian Artificial Intelligence Association

In cooperation with:

Japanese Society for Artificial Intelligence
Slovak Society for Informatics
The Institution of Electrical Engineers - Slovak Centre

Steering Committee:

Christo Dichev, IIT, Bulgarian Academy of Sciences Morio Nagata, Keio University Pavol Navrat, Slovak University of Technology Vadim L. Stefanuk, IITP, Russian Academy of Sciences Haruki Ueno, Tokyo Denki University

About the Conference

Joint Conference on Knowledge-Based Software Engineering aims to provide a forum for researchers and practitioners to discuss topics in knowledge engineering and in software engineering. Special emphasis is given to application of knowledge-based methods to software engineering problems. The conference originated from efforts to provide a suitable forum for contacts for scientists mainly from Japan, the CIS countries and the countries of the Central and Eastern Europe while always being open for participants from the whole world. JCKBSE'98 will continue in this tradition and expand for even greater international participation. Also, the scope of the conference as indicated by its topics is being updated to reflect the recent development in all three areas i.e.,

- knowledge engineering,

- software engineering,

- knowledge-based software engineering.

The conference will also include invited talks.

Topics (include, but are not limited to)

- Architecture of knowledge, software and information systems including collaborative, distributed, multi-agent and multimedia systems, internet and intranet

- Domain modelling

- Requirements engineering, formal and semiformal specifications

- Intelligent user interfaces and human machine interaction

- Knowledge acquisition and discovery, data mining

- Automating software design and synthesis

- Program understanding, programming knowledge

- Object-oriented and other programming paradigms, metaprogramming

- Reuse, re-engineering, reverse engineering

- Knowledge-based methods and tools for software engineering, including testing, verification and validation, process management, maintenance and evolution, CASE

- Decision support methods for software engineering

- Applied semiotics for knowledge-based software engineering

- Learning of programming, modelling programs and programmers

- Knowledge systems methodology, development tools and environments

- Software engineering and knowledge engineering education, distance learning, emergence of an information society

Program Committee

Zbigniew Banaszak Seiichi Komiya Technical Univ., Zielona Gora,PL IPA, Chair of SIG-KBSE, Co-chair, JPN Andras Benczur Behrouz H. Far Eotvos Lorand Univ.,Budapest,HU Saitama Univ., JPN Maria Bielikova Teruo Koyama Slovak Univ. of Technology, SK NACSIS, JPN Vladan Devedzic Vitaliy Lozovskiy Univ. of Belgrade, YU UAS, Odessa, UA Christo Dichev Ludovit Molnar IIT-BAS, BG Slovak U. of Technology, SK Darina Dicheva Morio Nagata Univ. of Sofia, BG Keio Univ., JPN Danail Dochev Pavol Navrat IIT-BAS, BG Slovak U. of Technology, Co-chair,SK Alexander Ehrlich Toshio Okamoto Computer Centre of RAS, RU U. of El. Communication, JPN Yoshiaki Fukazawa Gennadii Osipov Waseda Univ., JPN Programme Systems I. of RAS,RU Matjaz Gams Yury N. Pechersky Jozef Stefan I., Ljubljana, SI I. of Math. of MAS, MD Viktor Gladun Dmitrii Pospelov V.M.Glushkov I.,Kiev,UA Computer Centre of RAS, RU Vladimir Golenkov Vadim Stefanuk Radiotech. Univ. of Minsk,BY Inf.Transfer Probl.I.of RAS,RU Masaaki Hashimoto Kenji Sugawara Kyushu I. of Technol., JPN Chiba I. of Technology, JPN Tomas Hruska Enn Tyugu Technical Univ. of Brno, CZ Royal I. of Technology, Kista, SE Kenji Kaijiri Haruki Ueno Shinshu Univ., JPN Tokyo Denki Univ., JPN Vladimir Khoroshevsky Shuichiro Yamamoto Computer Centre of RAS, RU NTT, JPN

Venue

Smolenice castle is a beautiful site renown for providing excellent environment for scientific conferences. It serves as a congress centre. It is situated approximately 60km from Bratislava in the surroundings of Small Carpathians mountains.

Proceedings

All accepted papers will be published in the conference proceedings and will be available at the conference. In addition, several of the highest quality papers will be selected for a special issue of IEICE Transactions on Information and Systems.

Language

The official language of the conference will be English.

Paper Submission

Full papers should not exceed 8 pages. Short papers should not exceed 4 pages. Papers will be reviewed according to: technical quality, originality, clarity, appropriateness to the conference focus, and adequacy of references to related work. Authors should submit the papers electronically. For details see conference web site. In addition, one copy of a manuscript should be sent, too. Each paper should contain the following information:

- Title of the paper.

- Name, affiliation, mailing address of the author, e-mail.

- Abstract of 100-200 words.

- The subject category (the topic) in which the paper should be reviewed.

Important Dates

Feb. 1, 1998 - Registration forms
March 1, 1998 - Paper submission deadline
May 1, 1998 - Notification of acceptance
May 20, 1998 - Camera-ready deadline
Sept. 9 - 11, 1998 - Conference dates

Correspondence Address

JCKBSE'98
Department of Computer Science and Engineering
Slovak University of Technology
Ilkovicova 3, 812 19 Bratislava, Slovakia
fax: +421 7 720 415,
e-mail: jckbse98@dcs.elf.stuba.sk
    More information about JCKBSE'98 is available on the conference web site:
http://www.dcs.elf.stuba.sk/jckbse98/

Fee

Participants will pay for an integrated package comprising registration fee, board and lodging, proceedings and a half day excursion. Participants from CIS and Central and Eastern Europe are eligible for a substantially reduced fee.

Support For Students

There will be available a limited number of scholarships for students submitting papers to the conference to support partially their participation.

Tenth IASTED International Conference on Parallel and Distributed Computing and Systems

Las Vegas, Nevada, U.S.A., October 28-31, 1998
Sponsored by International Association of Science and Technology for Development (IASTED)
http://www.cps.udayton.edu/~pan/pdcs98

Purpose

The International Conference on Parallel and Distributed Computing and systems, sponsored by IASTED, is a major annual forum for scientists, engineers, and practitioners throughout the world to present the latest research results, ideas, development, and applications in all areas of parallel and distributed processing. The 1997 conference attracted researchers from 30 countries. The 1998 meeting (PDCS '98) will be held in Las Vegas, Nevada, U.S.A., and will include keynote addresses, contributed papers, tutorials, and workshops.

Scope

The main focus of PDCS '98 will be parallel and distributed computing and systems viewed from the three perspectives of architecture and networking, software systems, and algorithms and applications. Topics include, but are not limited to, the following:

Architecture and Networking: SIMD/MIMD processors, various parallel/concurrent architecture styles, interconnection networks, memory systems and management, I/O in parallel processing, VLSI systems, optical computing, computer networks, communications and telecommunications, wireless networks and mobile computing.

Software Systems: operating systems, programming languages, various parallel programming paradigms, vectorization and program transformation, parallelizing compilers, tools and environments for software development, distributed data- and knowledge-base systems, modelling and simulation, performance evaluation and measurements, visualization.

Algorithms and Applications: parallel/distributed algorithms, resource allocation and management, load sharing and balancing, task mapping and job scheduling, network routing and communication algorithms, reliability and fault tolerance, signal and image processing, neural networks, high-performance scientific computing, application studies.

Paper Submission Guidelines

Papers reporting original and unpublished research results and experience are solicited. Papers will be selected based on their originality, timeliness, significance, relevance, and clarity of presentation. Accepted and presented papers will be published in the conference proceedings.

Please send four copies of a manuscript to the program committee chair at the following address by April 15, 1998: Professor Yi Pan, PDCS '98 Program Chair, Department of Computer Science, University of Dayton, Dayton, Ohio, 45469-2160, U.S.A. Phone: (937) 229-3807. Fax: (937) 229-4000. Email: pan@udcps.cps.udayton.edu.

A manuscript should not exceed 15 pages, including tables and figures. In the cover letter, please indicate the author for correspondence, and his/her complete postal address, phone and fax numbers, and email address (make sure the email address is current and working).

Tutorials/Workshops

Several workshops are being planned for PDCS '98. Each workshop will focus on a particular topic, and consists of several presentations and open discussion. A one-page abstract of each workshop presentation will be published in the conference proceedings. The proposal for a workshop should include the title, topics covered, proposed/invited speakers, and estimated length (hours) of the workshop. Anyone wishing to organize a workshop in connection with PDCS '98 should submit four copies of his/her proposal to the program chair at the address given above by April 15, 1998.

PDCS '98 will also offer half-day tutorials in parallel and distributed computing. Each tutorial proposal should provide the title, topics, targeted audiences, and instructor's biography. The proposal should be submitted to the tutorial chair Dr. Pradip K. Srimani via email at srimani@CS.Colostate.EDU.

IJPDSN

A special issue consisting of selected papers from PDCS '98 will be published in International Journal of Parallel and Distributed Systems and Networks. Other papers may be submitted to: Professor Marlin H. Mickle, Editor-In-Chief, IJPDSN, Department of

Electrical Engineering, University of Pittsburgh, Pittsburgh, PA 15261, USA.

Important Dates

Paper submission deadline: **April 15, 1998**
Author notification: June 15, 1998
Camera-ready version due: August 1, 1998
Workshop/tutorial proposal due: April 15, 1998
Conference: October 28-31, 1998

General Co-Chairs

Selim G. Akl, Queens University (Canada)
Keqin Li, State University of New York (USA)

Program Chair

Yi Pan, University of Dayton (USA)

Tutorial Chair

Pradip K. Srimani, Colorado State University (USA)

Local Arrangement Chair

Shahram Latifi, University of Nevada at Las Vegas (USA)

Program Committee

Ishfaq Ahmad, Hong Kong Univ. of Sci. & Tech. (H. K.)
Hamid R. Arabnia, University of Georgia (USA)
Mohammed Atiquzzaman, University of Dayton (USA)
Johnnie W. Baker, Kent State University (USA)
David A. Berson, Intel Corporation (USA)
Jingde Cheng, Kyushu University (Japan)
Kam-Hoi Cheng, University of Houston (USA)
Henry Chuang, University of Pittsburgh (USA)
Kuo-Liang Chung, Nat'l Taiwan U. of Sci. & Tech (Taiwan)
Bin Cong, California Polytechnic State University (USA)
Mark Cross, University of Greenwich (UK)
Sajal K. Das, University of North Texas (USA)
Brian J. d'Auriol, The University of Akron (USA)
Frank Dehne, Carleton University (Canada)
Eliezer Dekel, IBM - Haifa Research Lab. (Israel)
Ivan Dimov, Bulgarian Academy of Sciences (Bulgaria)
Erik H. D'Hollander, University of Ghent (Belgium)
Omer Egecioglu, Univ. of California at Santa Barbara (USA)
Hossam ElGindy, University of Newcastle (Australia)
Afonso Ferreira, CNRS-INRIA (France)
Rajiv Gupta, University of Pittsburgh (USA)
Mounir Hamdi, Hong Kong Univ. of Sci. & Tech. (H. K.)
Jack Jean, Wright State University (USA)
Weijia Jia, City University of Hong Kong (Hong Kong)
Shahram Latifi, University of Nevada (USA)
Yamin Li, University of Aizu (Japan)
Ahmed Louri, University of Arizona (USA)
Bruce Maggs, Carnegie Mellon University (USA)
Brian Malloy, Clemson University (USA)
Koji Nakano, Nagoya Institute of Technology (Japan)

William T. O'Connell, Lucent Bell Labs (USA)
Michael A. Palis, Rutgers University (USA)
Marcin Paprzycki, University of Southern Mississippi (USA)
Behrooz Parhami, Univ. of Calif. at Santa Barbara (USA)
Chan-Ik Park, Pohang Univ. of Science & Tech (Korea)
Lori Pollock, University of Delaware (USA)
Jerry L. Potter, Kent State University (USA)
Chunming Qiao, State University of New York (USA)
C. S. Raghavendra, The Aerospace Corporation (USA)
Sanguthevar Rajasekaran, University of Florida (USA)
Hesham El-Rewini, University of Nebraska at Omaha (USA)
Jose D. P. Rolim, University of Geneva (Switzerland)
Hemant Rotithor, Digital Equipment Corporation (USA)
Ponnuswamy Sadayappan, Ohio State University (USA)
Assaf Schuster, Technion (Israel)
Edwin Sha, University of Notre Dame (USA)
Hong Shen, Griffith University (Australia)
Xiaojun Shen, University of Missouri at Kansas City (USA)
Gurdip Singh, Kansas State University (USA)
Mary Lou Soffa, University of Pittsburgh (USA)
Pradip K. Srimani, Colorado State University (USA)
Per Stenstrom, Chalmers Univ. of Technology (Sweden)
Ivan Stojmenović, University of Ottawa (Canada)
Hal Sudborough, University of Texas at Dallas (USA)
Jerry L. Trahan, Louisiana State University (USA)
Ramachandran Vaidyanathan, Louisiana State Univ. (USA)
Subbarayan Venkatesan, University of Texas at Dallas (USA)
Alan S. Wagner, University of British Columbia (Canada)
Yuanyuan Yang, University of Vermont (USA)

Steering Committee

Narsingh Deo, University of Central Florida (USA)
Ahmed Elmagarmid, Purdue University (USA)
Geoffrey Fox, Syracuse University (USA)
H. Scott Hinton, University of Colorado (USA)
D. Frank Hsu, Fordham University (USA)
Oscar Ibarra, Univ. of California at Santa Barbara (USA)
Joseph JáJá, University of Maryland (USA)
Lennart Johnsson, University of Houston (USA)
H. T. Kung, Harvard University (USA)
Tom Leighton, Massachusetts Inst. of Technology (USA)
Jane Liu, University of Illinois at Urbana Champaign (USA)
Marlin Mickle, University of Pittsburgh (USA)
Lionel Ni, Michigan State University (USA)
Stephan Olariu, Old Dominion University (USA)
Sartaj Sahni, University of Florida (USA)
Eugen Schenfeld, NEC Research Institute (USA)
Marc Snir, IBM - Thomas Watson Research Center (USA)
Hal Sudborough, University of Texas at Dallas (USA)
Si-Qing Zheng, Louisiana State University (USA)
Albert Y. Zomaya, Univ. of Western Australia (Australia)

THE MINISTRY OF SCIENCE AND TECHNOLOGY OF THE REPUBLIC OF SLOVENIA

Address: Slovenska 50, 1000 Ljubljana, Tel.: +386 61 1311 107, Fax: +386 61 1324 140.
WWW:http://www.mzt.si
Minister: Lojze Marinček, Ph.D.

The Ministry also includes:

The Standards and Metrology Institute of the Republic of Slovenia
Address: Kotnikova 6, 61000 Ljubljana, Tel.: +386 61 1312 322, Fax: +386 61 314 882.

Slovenian Intellectual Property Office
Address: Kotnikova 6, 61000 Ljubljana, Tel.: +386 61 1312 322, Fax: +386 61 318 983.

Office of the Slovenian National Commission for UNESCO
Address: Slovenska 50, 1000 Ljubljana, Tel.: +386 61 1311 107, Fax: +386 61 302 951.

Scientific, Research and Development Potential:

The Ministry of Science and Technology is responsible for the R&D policy in Slovenia, and for controlling the government R&D budget in compliance with the National Research Program and Law on Research Activities in Slovenia. The Ministry finances or co-finance research projects through public bidding, while it directly finance some fixed cost of the national research institutes.

According to the statistics, based on OECD (Frascati) standards, national expenditures on R&D raised from 1,6 % of GDP in 1994 to 1,71 % in 1995. Table 2 shows an income of R&D organisation in million USD.

Objectives of R&D policy in Slovenia:

- maintaining the high level and quality of scientific technological research activities;

- stimulation and support to collaboration between research organisations and business, public, and other sectors;

Total investments in R&D (% of GDP)	1,71
Number of R&D Organisations	297
Total number of employees in R&D	12.416
Number of researchers	6.094
Number of Ph.D.	2.155
Number of M.Sc.	1.527

Table 1: Some R&D indicators for 1995

	Ph.D.			M.Sc.		
	1993	1994	1995	1993	1994	1995
Bus. Ent.	51	93	102	196	327	330
Gov. Inst.	482	574	568	395	471	463
Priv. np Org.	10	14	24	12	25	23
High. Edu.	1022	1307	1461	426	772	711
TOTAL	1565	1988	2155	1029	1595	1527

Table 2: Number of employees with Ph.D. and M.Sc.

- stimulating and supporting of scientific and research disciplines that are relevant to Slovenian national authenticity;

- co-financing and tax exemption to enterprises engaged in technical development and other applied research projects;

- support to human resources development with emphasis on young researchers; involvement in international research and development projects;

- transfer of knowledge, technology and research achievements into all spheres of Slovenian society.

Table source: Slovene Statistical Office.

	Basic Research		Applied Research		Exp. Devel.		Total	
	1994	1995	1994	1995	1994	1995	1994	1995
Business Enterprises	6,6	9,7	48,8	62,4	45,8	49,6	101,3	121,7
Government Institutes	22,4	18,6	13,7	14,3	9.9	6,7	46,1	39,6
Private non-profit Organisations	0,3	0,7	0,9	0,8	0,2	0,2	1,4	1,7
Higher Education	17,4	24,4	13,7	17,4	8,0	5,7	39,1	47,5
TOTAL	46,9	53,4	77,1	94,9	63.9	62,2	187,9	210,5

Table 3: Incomes of R&D organisations by sectors in 1995 (in million USD)

JOŽEF STEFAN INSTITUTE

Jožef Stefan (1835-1893) was one of the most prominent physicists of the 19th century. Born to Slovene parents, he obtained his Ph.D. at Vienna University, where he was later Director of the Physics Institute, Vice-President of the Vienna Academy of Sciences and a member of several scientific institutions in Europe. Stefan explored many areas in hydrodynamics, optics, acoustics, electricity, magnetism and the kinetic theory of gases. Among other things, he originated the law that the total radiation from a black body is proportional to the 4th power of its absolute temperature, known as the Stefan–Boltzmann law.

The Jožef Stefan Institute (JSI) is the leading independent scientific research institution in Slovenia, covering a broad spectrum of fundamental and applied research in the fields of physics, chemistry and biochemistry, electronics and information science, nuclear science technology, energy research and environmental science.

The Jožef Stefan Institute (JSI) is a research organisation for pure and applied research in the natural sciences and technology. Both are closely interconnected in research departments composed of different task teams. Emphasis in basic research is given to the development and education of young scientists, while applied research and development serve for the transfer of advanced knowledge, contributing to the development of the national economy and society in general.

At present the Institute, with a total of about 700 staff, has 500 researchers, about 250 of whom are postgraduates, over 200 of whom have doctorates (Ph.D.), and around 150 of whom have permanent professorships or temporary teaching assignments at the Universities.

In view of its activities and status, the JSI plays the role of a national institute, complementing the role of the universities and bridging the gap between basic science and applications.

Research at the JSI includes the following major fields: physics; chemistry; electronics, informatics and computer sciences; biochemistry; ecology; reactor technology; applied mathematics. Most of the activities are more or less closely connected to information sciences, in particular computer sciences, artificial intelligence, language and speech technologies, computer-aided design, computer architectures, biocybernetics and robotics, computer automation and control, professional electronics, digital communications and networks, and applied mathematics.

The Institute is located in Ljubljana, the capital of the independent state of **Slovenia** (or S♡nia). The capital today is considered a crossroad between East, West and Mediterranean Europe, offering excellent productive capabilities and solid business opportunities, with strong international connections. Ljubljana is connected to important centers such as Prague, Budapest, Vienna, Zagreb, Milan, Rome, Monaco, Nice, Bern and Munich, all within a radius of 600 km.

In the last year on the site of the Jožef Stefan Institute, the Technology park "Ljubljana" has been proposed as part of the national strategy for technological development to foster synergies between research and industry, to promote joint ventures between university bodies, research institutes and innovative industry, to act as an incubator for high-tech initiatives and to accelerate the development cycle of innovative products.

At the present time, part of the Institute is being reorganized into several high-tech units supported by and connected within the Technology park at the Jožef Stefan Institute, established as the beginning of a regional Technology park "Ljubljana". The project is being developed at a particularly historical moment, characterized by the process of state reorganisation, privatisation and private initiative. The national Technology Park will take the form of a shareholding company and will host an independent venture-capital institution.

The promoters and operational entities of the project are the Republic of Slovenia, Ministry of Science and Technology and the Jožef Stefan Institute. The framework of the operation also includes the University of Ljubljana, the National Institute of Chemistry, the Institute for Electronics and Vacuum Technology and the Institute for Materials and Construction Research among others. In addition, the project is supported by the Ministry of Economic Relations and Development, the National Chamber of Economy and the City of Ljubljana.

Jožef Stefan Institute
Jamova 39, 61000 Ljubljana, Slovenia
Tel.:+386 61 1773 900, Fax.:+386 61 219 385
Tlx.:31 296 JOSTIN SI
WWW: http://www.ijs.si
E-mail: matjaz.gams@ijs.si
Contact person for the Park: Iztok Lesjak, M.Sc.
Public relations: Natalija Polenec

INFORMATICA

AN INTERNATIONAL JOURNAL OF COMPUTING AND INFORMATICS

INVITATION, COOPERATION

Submissions and Refereeing

Please submit three copies of the manuscript with good copies of the figures and photographs to one of the editors from the Editorial Board or to the Contact Person. At least two referees outside the author's country will examine it, and they are invited to make as many remarks as possible directly on the manuscript, from typing errors to global philosophical disagreements. The chosen editor will send the author copies with remarks. If the paper is accepted, the editor will also send copies to the Contact Person. The Executive Board will inform the author that the paper has been accepted, in which case it will be published within one year of receipt of e-mails with the text in Informatica LaTeX format and figures in .eps format. The original figures can also be sent on separate sheets. Style and examples of papers can be obtained by e-mail from the Contact Person or from FTP or WWW (see the last page of Informatica).

Opinions, news, calls for conferences, calls for papers, etc. should be sent directly to the Contact Person.

QUESTIONNAIRE

☐ Send Informatica free of charge

☐ Yes, we subscribe

Please, complete the order form and send it to Dr. Rudi Murn, Informatica, Institut Jožef Stefan, Jamova 39, 61111 Ljubljana, Slovenia.

Since 1977, Informatica has been a major Slovenian scientific journal of computing and informatics, including telecommunications, automation and other related areas. In its 16th year (more than five years ago) it became truly international, although it still remains connected to Central Europe. The basic aim of Informatica is to impose intellectual values (science, engineering) in a distributed organisation.

Informatica is a journal primarily covering the European computer science and informatics community - scientific and educational as well as technical, commercial and industrial. Its basic aim is to enhance communications between different European structures on the basis of equal rights and international refereeing. It publishes scientific papers accepted by at least two referees outside the author's country. In addition, it contains information about conferences, opinions, critical examinations of existing publications and news. Finally, major practical achievements and innovations in the computer and information industry are presented through commercial publications as well as through independent evaluations.

Editing and refereeing are distributed. Each editor can conduct the refereeing process by appointing two new referees or referees from the Board of Referees or Editorial Board. Referees should not be from the author's country. If new referees are appointed, their names will appear in the Refereeing Board.

Informatica is free of charge for major scientific, educational and governmental institutions. Others should subscribe (see the last page of Informatica).

ORDER FORM – INFORMATICA

Name: ....

Title and Profession (optional): ....

....

Home Address and Telephone (optional): ....

....

Office Address and Telephone (optional): ....

....

E-mail Address (optional): ....

Signature and Date: ....

Referees:

EDITORIAL BOARDS, PUBLISHING COUNCIL

Informatica

An International Journal of Computing and Informatics

Contents: